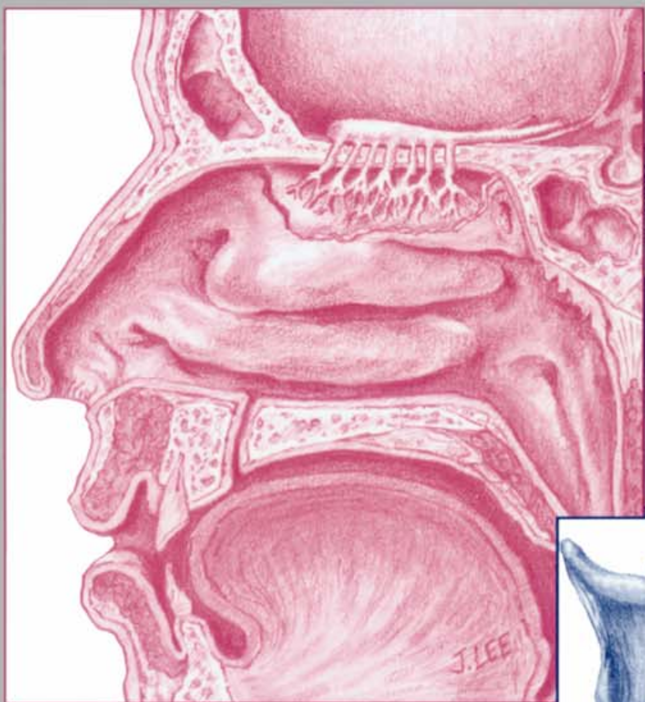
Otolaryngology

Basic Science and Clinical Review

Thomas R. Van De Water
Hinrich Staecker



Thieme

Otolaryngology

Basic Science and Clinical Review

Thieme

This page intentionally left blank

OTOLARYNGOLOGY

Basic Science and Clinical Review

EDITOR
Thomas R. Van De Water, Ph.D.
Director
Cochlear Implant Research Program
University of Miami Ear Institute
Professor
Department of Otolaryngology
Miller School of Medicine
University of Miami
Miami, Florida

ASSOCIATE EDITOR
Hinrich Staecker, M.D., Ph.D.
Director
Otology and Neurotology Program
Associate Professor
Department of Otolaryngology–Head and Neck Surgery
University of Maryland School of Medicine
Baltimore, Maryland

Thieme
New York • Stuttgart

Thieme Medical Publishers, Inc.
333 Seventh Ave.
New York, NY 10001

Editor: Esther Gumpert
Associate Editor: Birgitta Brandenburg
Vice President, Production and Electronic Publishing: Anne T. Vinnicombe
Production Editor: Print Matters, Inc.
Sales Director: Ross Lumpkin
Associate Marketing Director: Verena Diem
Chief Financial Officer: Peter van Woerden
President: Brian D. Scanlan
Compositor: Compset
Printer: Maple-Vail Book Manufacturing Group

Library of Congress Cataloging-in-Publication Data
Otolaryngology : basic science and clinical review / [edited by] Thomas R. Van De Water,
Hinrich Staecker.
p. ; cm.
Includes bibliographical references and index.
ISBN 0-86577-901-5 (US)—ISBN 3-13-124651-0 (GTV)
1. Otolaryngology. 2. Ear—Physiology. 3. Respiratory organs—Physiology.
[DNLM: 1. Otorhinolaryngologic Diseases. 2. Ear—physiology. 3. Otorhinolaryngologic
Surgical Procedures. 4. Respiratory Physiology. WV 150 088 2005] I. Van De Water,
Thomas R. II. Staecker, Hinrich.
RF46.07525 2005
616.5'1—dc22 2005050640

Important note: Medical knowledge is ever-changing. As new research and clinical experience broaden our knowledge, changes in treatment and drug therapy may be required. The authors and editors of the material herein have consulted sources believed to be reliable in their efforts to provide information that is complete and in accord with the standards accepted at the time of publication. However, in view of the possibility of human error by the authors, editors, or publisher of the work herein or changes in medical knowledge, neither the authors, editors, or publisher, nor any other party who has been involved in the preparation of this work, warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from use of such information. Readers are encouraged to confirm the information contained herein with other sources. For example, readers are advised to check the product information sheet included in the package of each drug they plan to administer to be certain that the information contained in this publication is accurate and that changes have not been made in the recommended dose or in the contraindications for administration. This recommendation is of particular importance in connection with new or infrequently used drugs.

Printed in the United States of America
5 4 3 2 1
TMP ISBN 0-86577-901-5
GTV ISBN 3 13 124651 0

Maxwell Abramson, M.D. (1935–1991)

This book is dedicated to the memory of Maxwell Abramson, M.D., husband, father, physician, educator, researcher, scholar, and friend. Max was an extraordinary human being whose life reflected his core humanitarian values. He was a gifted healer, creative scientist, and talented teacher in the discipline of otolaryngology. He possessed both a strong desire and an unbridled enthusiasm to pass along his knowledge of the basic science and clinical foundations of otolaryngology to the fellows, residents, and medical students at the Columbia University College of Physicians and Surgeons, where he was chairman of the otolaryngology department from 1977 until his untimely death in 1991. Max greatly loved both his family and his chosen profession. He was a valued friend to many of us in the otolaryngology community, and his absence is felt by all of us. I know that Max would be pleased to have this book dedicated to his memory because it continues the dissemination of basic and clinical sciences knowledge of otolaryngology.

Thomas R. Van De Water, Ph.D.

Thispageintentionallyleftblank

Contents

Preface

For otolaryngology–head and neck surgeons in training as residents and fellows, it is a difficult task to keep up with all of the advances in the basic and clinical sciences that impact the medical and surgical practice of their specialty. It is also pertinent to the practice of otolaryngology–head and neck surgery for interested medical students and lecturers in medical school faculties who teach students and resident physicians to have a reference book that clearly presents the basic principles. This book serves as an excellent resource for residents preparing for board exams, students studying for exams, and a refresher for practioners and all other interested parties. It is a natural by-product of a Basic Sciences in Otolaryngology course that was taught at the New York Academy of Medicine. This course was first organized as a Saturday teaching program, and at the behest of Professors Max Abramson (chairman of the Department of Otolaryngology, College of Physicians and Surgeons, Columbia University) and Bob Ruben (chairman of the Department of Otolaryngology, Albert Einstein College of Medicine, Yeshiva University), I reorganized and expanded the scope of the course and became the academic director for over 15 years. Six New York and one New Jersey otolaryngology–head and neck surgery residency training programs participated in this basic sciences course on Tuesday evenings at the New York Academy of Medicine. The course became very comprehensive and covered almost all of the basic and clinical science aspects of otolaryngology–head and neck surgery. In 1991, I renamed this resident teaching program the Maxwell Abramson Basic Sciences Course in Otolaryngology to honor Max after his untimely and tragic death. The content and relevancy of this course have been continually improved by responding to the critiques and suggestions provided by the attending residents, medical students, fellows, residency program directors, program chairs, and lecturers. It remained current by the hard work of the invited lecturers who continually incorporated contemporary advances into the talks they gave that covered their respective fields of expertise. The subject matter and information presented in this book are a product of that process of updating and refinement.

Because this book attempts to cover the entire field of otolaryngology–head and neck surgery, it is very comprehensive and therefore is composed of 60 chapters. These chapters are arranged into six broadly defined sections. Chapters 1 to 19 comprise Section I, The Basic Principles; Chapters 20 to 36 form Section II, The Ear, Hearing, and Balance; Chapters 37 to 40 comprise Section III, The Nose, Olfaction, and the Sinuses; chapters 41 to 51 cover section IV, The Larynx, Voice, and Neck; Chapters 52 to 55 represent section V, The Oral Cavity, Taste, and the Glands of the Neck; and Chapters 56 to 60 are Section VI, Facial Plastics and Miscellaneous.

The authors invited to write each chapter were selected because they are leaders and experts in their subject areas and for their ability to confer their knowledge in a clear and concise format that is appropriate for the targeted audience. There are over 100 authors who have contributed to *Otolaryngology: Basic Science and Clinical Review*. These authors come from such diverse fields as anatomy, auditory physiology, pharmacology, radiology, general otolaryngology, molecular biology, molecular genetics, plastic surgery, infectious diseases, pediatrics, otology, neurotology, phonosurgery, voice, head and neck surgery, and oncology. They hold such diverse degrees as Ph.D.s, Au.D.s, and M.D.s, but a uniting factor is that they are all teachers with a desire to educate, who consider it both their duty and a privilege to pass on their accumulated knowledge to interested students, residents, and fellows.

This book is dedicated not only to the memory of Professor Maxwell Abramson, who was both a caring physician and a gifted teacher, but also to the many residents, medical students, and fellows who have participated in this course over the years and whose intelligent and thoughtful input improved the content of the Maxwell

Abramson Basic Sciences in Otolaryngology course and therefore the final content of this book. We also thank the many authors who spent countless hours writing, rewriting, and updating their chapters, for without their dedication and hard work, none of this would have been possible.

No book is accomplished without many hours that turn into days and then into weeks and months stolen from our families with their knowledge and assent. We acknowledge this sacrifice and express our deep gratitude to our wives, Jeanette Van De Water and Danielle Staecker, for both their understanding and generosity of spirit.

Otolaryngology: Basic Science and Clinical Review could never have seen the light of day without the very able assistance and help from our editor, Esther Gumpert, and our very capable, committed, and multitalented associate editor, Birgitta Brandenburg, at Thieme Medical Publishers in New York.

Thomas R. Van De Water, Ph.D.
Hinrich Staecker, M.D., Ph.D.

Foreword

Otolaryngology: Basic Science and Clinical Review fills a unique requirement for a contemporary, definitive textbook that covers the expansive fields of basic and clinical sciences in otolaryngology.

The education of medical students and residents to become practicing otolaryngologists is a primary function of any department of otolaryngology–head and neck surgery. To become skilled clinicians and surgeons, residents must have a firm understanding of the scientific precepts that form the bases for the clinical and surgical practice of their chosen specialty.

This textbook is the natural outgrowth of a basic science in otolaryngology course that was organized and taught by Thomas Van De Water, Ph.D., at the New York Academy of Medicine for over 15 years. The textbook benefits from refinements and the continuous updating of that course in response to feedback from students, residents, and faculty. Knowledge is presented in a clear and comprehensive format written by experts in each of the disciplines who have kept their targeted audience in mind.

There is at present no other textbook in the specialty of otolaryngology–head and neck surgery that brings together all of the basic and clinical science knowledge needed for a comprehensive understanding of this surgical/medical specialty. As the first of its genre, *Otolaryngology: Basic Science and Clinical Review* fills a much needed place in the education of residents, provides a resource for medical students interested in pursuing a career in this specialty, and acts as a study guide for recent graduates of training programs. This ambitious text—detailing the explosion of knowledge underpinning a highly diverse specialty—is destined to become both required reading for residents and an authoritative reference for practicing otolaryngologists. Professor Van De Water and associate editor Hinrich Staecker, M.D., Ph.D., along with the expert authors gathered here, are to be congratulated for this outstanding contribution to our specialty.

Thomas J. Balkany, M.D., F.A.C.S., F.A.A.P.

Thispageintentionallyleftblank

Contributors

EDITOR

Thomas R. Van De Water, Ph.D.
Director
Cochlear Implant Research Program
University of Miami Ear Institute
Professor
Department of Otolaryngology
Miller School of Medicine
University of Miami
Miami, Florida

ASSOCIATE EDITOR

Hinrich Staecker, M.D., Ph.D.
Director
Otology and Neurotology Program
Associate Professor
Department of Otorhinolaryngology–Head and Neck Surgery
University of Maryland School of Medicine
Baltimore, Maryland

CONTRIBUTORS

Simon I. Angeli, M.D.
Associate Professor
Department of Otolaryngology
Miller School of Medicine
University of Miami
Miami, Florida

Abigail Arad-Cohen, M.D.
Shoham, Israel

Armand Balboni, J.D., M.Phil. (A.B.D.)
Department of Anatomy and Morphology
Mount Sinai School of Medicine
New York, New York

Thomas J. Balkany, M.D., F.A.C.S., F.A.A.P.
Hotchkiss Professor and Chairman
Department of Otolaryngology
Professor of Neurological Surgery and Pediatrics
Miller School of Medicine
University of Miami
Miami, Florida

Soly Baredes, M.D., F.A.C.S.
Chairman
Department of Otolaryngology
University of Medicine and Dentistry of New Jersey
Newark, New Jersey

Andrew Bergemann, Ph.D.
Department Of Pathology
Mount Sinai School of Medicine
New York, New York

Diana V.-A. Berggren, M.D., Ph.D.
Chairman
Department of Otolaryngology
Professor
Departments of Clinical Science, Otorhinolaryngology
Umeå University
Umeå, Sweden

Andrew Blitzer, M.D., D.D.S.
Professor
Department of Otolaryngology
College of Physicians & Surgeons
Columbia University
Presbyterian Hospital
New York, New York

K. Paul Boyev, M.D.
Associate Professor
Department Of Otolaryngology
University of South Florida
Tampa, Florida

Marjorie S Brandwein-Gensler, M.D.
Professor
Departments of Pathology and Otolaryngology
Mount Sinai School of Medicine
Mount Sinai Medical Center
New York, New York

Kathleen Campbell, Ph.D.
Professor and Director of Audiology Research
Department of Surgery
Southern Illinois University School of Medicine
Springfield, Illinois

John Carey, M.D.
Associate Professor
Department of Otolaryngology–Head and Neck Surgery
The Johns Hopkins University School of Medicine
Baltimore, Maryland

Lanny Garth Close, M.D.
Howard W. Smith Professor and Chair
Department of Otolaryngology–Head and Neck Surgery
College of Physicians & Surgeons
Columbia University
Presbyterian Hospital
New York, New York

Noel L. Cohen, M.D.
Chairman Emeritus
Department of Otolaryngology
New York University School of Medicine
New York, New York

John K. M. Coleman, Ph.D.
Senior Research Scientist
Department of Otolaryngology
Naval Medical Center–San Diego
San Diego, California

Robin T. Cotton, M.D.
Professor
Department of Otolaryngology
Children's Hospital Medical Center
University of Cincinnati College of Medicine
Cincinnati, Ohio

David R. Edelstein, M.D., F.A.C.S.
Chief
Division of Otolaryngology/Nasal Sinus Disease/Endoscopy
Manhattan Eye, Ear and Throat Hospital
New York, New York

Adrien A. Eshraghi, M.Sc., M.D.
Associate Professor
Department of Otolaryngology
Miller School of Medicine
University of Miami
Jackson Memorial Hospital
Miami, Florida

Joseph Feghali, M.D., F.A.C.S.
Adjunct Professor
Department of Otolaryngology and Neurological Surgery
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

Elizabeth J. Franzmann, M.D.
Assistant Professor
Department of Otolaryngology
Miller School of Medicine
University of Miami
Miami, Florida

Dorothy Frenz, Ph.D.
Director
Otolaryngology Research
Professor
Departments of Otolaryngology and Anatomy and Structural Biology
Albert Einstein College of Medicine
Bronx, New York

Neal Futran, M.D.
Assistant Professor
Department of Otolaryngology–Head and Neck Surgery
University of Washington
Seattle, Washington

Peter G. Gillespie, Ph.D.
Professor of Otolaryngology and Cell Biology
Oregon Hearing Research Center
Oregon Health and Sciences University
Portland, Oregon

Bradley J. Goldstein, M.D., Ph.D.
Assistant Professor
Department of Otolaryngology–Head and Neck Surgery
The Johns Hopkins University School of Medicine
Baltimore, Maryland

Alexander W. Gotta, M.D.
Chairman Emeritus
Department of Pharmacology
Downstate Medical Center
State University of New York
Brooklyn, New York

John H. Greinwald, Jr., M.D.
Associate Professor of Otolaryngology and Pediatrics
Department of Otolaryngology
Division of Pediatric Otolaryngology
Children's Hospital Medical Center
University of Cinncinati College of Medicine
Cincinnati, Ohio

John H. Grose, M.D.
Department of Otolaryngology–Head and Neck Surgery
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina

Jeffrey P. Harris, M.D.
Chairman
Department of Otolaryngology
Head and Neck Surgery Clinic
University of California–San Diego Medical Center
La Jolla, California

Christopher J. Hartnick, M.D.
Assistant Professor
Department of Otology and Laryngology
Harvard Medical School
Massachusetts Eye and Ear Infirmary
Boston, Massachusetts

Gerald B. Healy, M.D.
Professor
Department of Otology and Laryngology
Children's Hospital
Harvard Medical School
Boston, Massachusetts

G. Richard Holt, M.D., M.S.E., M.P.H.
University of Texas Health Science Center
San Antonio, Texas

David Huang, M.D., Ph.D.
Department of Radiation Oncology
Miller School of Medicine
University of Miami
Miami, Florida

Golda Hudes, M.D., Ph.D.
Assistant Professor
Department of Medicine (Allergy & Immunology)
Department of Otolaryngology
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

Steven R. Issacson, M.D., F.A.C.S.
Associate Professor
Departments of Radiation Oncology and Otolaryngology
College of Physicians & Surgeons
Columbia University
Presbyterian Hospital
New York, New York

Ronald Jackson, Ph.D.
Senior Scientist
Department of Otolaryngology
Naval Medical Center
San Diego, California

R. L. Jackson, M.D.
Department of Defense Spatial Orientation Center
Naval Medical Center
San Diego, California

Brian Jewett, M.D.
Assistant Professor
Department of Otolaryngology
Division of Facial Plastic Surgery
Miller School of Medicine
University of Miami
Miami, Florida

Charles P. Kimmelmann, M.D., M.B.A., F.A.C.S.
Associate Clinical Professor
Department of Otorhinolaryngology
Weill Cornell Medical Center
New York, New York

Richard D. Kopke, M.D., F.A.C.S.
Clinical Professor
Department of Otorhinolaryngology
University of Oklahoma Health Sciences Center
Director, Hough Ear Institute
Oklahoma City, Oklahoma

Lane Krevitt, M.D.
Adjunct Clinical Instructor
Department of Otolaryngology
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

Daniel B. Kuriloff, M.D., F.A.C.S.
Associate Professor
Department of Otolaryngology–Head and Neck Surgery
College of Physicians & Surgeons
Columbia University
Presbyterian Hospital
New York, New York

Daniel B. Kurtz, Ph.D.
Assistant Professor
Department of Biology
Utica College
Utica, New York

Jeffrey T. Laitman, Ph.D.
Professor and Director
Center for Anatomy and Functional Morphology
Professor, Department of Otolaryngology
Mount Sinai School of Medicine
New York, New York

Philippe P. Lefebvre, M.D., Ph.D.
Professor and Chairman
Department of Otolaryngology and Audiophonology
University of Liège
Liège, Belgium

Alan David Legatt, M.D., Ph.D.
Professor
Department of Neurology
Albert Einstein College of Medicine
Director, EEG Laboratory
Director, Evoked Potential Laboratory
Director, Intraoperative Neurophysiology
Montefiore Medical Center
Bronx, New York

Ira M. Leviton, M.D.
Department of Internal Medicine
Division of Infectious Diseases
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

M. Charles Liberman, Ph.D.
Professor
Department of Otology and Laryngology
Harvard Medical School
Director
Eaton Peabody Laboratory
Massachusetts Eye and Ear Infirmary
Boston, Massachusetts

Scott Lilly, D.D.S., M.D.
Oklahoma Oncology, Inc.
Tulsa, Oklahoma

David J. Lim, M.D.
Executive Vice President of Research
Head, Gonda Department of Cell and Molecular Biology
House Ear Institute
Adjunct Professor
University of Southern California
Los Angeles, California

Jianzhong Liu, M.D.
Department of Otorhinolaryngology
University of Oklahoma Health Sciences Center
Hough Ear Institute
Oklahoma City, Oklahoma

Dinesh Mehta, M.D., F.A.C.S., F.R.C.S.
Clinical Associate Professor
Department of Otolaryngology–Head and Neck Surgery
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

Saumil N. Merchant, M.D.
Associate Professor
Department of Otology and Laryngology
Harvard Medical School
Massachusetts Eye and Ear Infirmary
Boston, Massachusetts

Kristine Mosier, D.M.D., Ph.D.
Assistant Professor
Department of Radiology
Indiana University School of Medicine
Indianapolis, Indiana

Joseph B. Nadol, Jr., M.D.
Professor and Chairman
Department of Otology and Laryngology
Harvard Medical School
Massachusetts Eye and Ear Infirmary
Boston, Massachusetts

Hari Nadiminti, M.D.
Resident, Internal Medicine
Harvard Medical School
Massachusetts General Hospital
Boston, Massachusetts

Drew M. Noden, Ph.D.
Professor of Embryology and Animal Development
Department of Biomedical Sciences
College of Veterinary Medicine
Cornell University
Ithaca, New York

Jane A. Petro, M.D., F.A.C.S.
Professor
Department of Surgery
Division of Plastic Surgery
New York Medical College
White Plains, New York

Walter M. Ralph, Jr., M.D., Ph.D.
Department of Surgery
Division of Ololaryngology
St Johns Queens Hospital
Elmhurst, New York

Gregory W. Randolph, M.D.
Department of Otology and Laryngology
Harvard Medical School
Massachusetts Eye and Ear Infirmary
Boston, Massachusetts

Steven D. Rauch, M.D.
Associate Professor
Department of Otolaryngology
Harvard Medical School
Massachusetts Eye and Ear Infirmary
Boston, Massachusetts

Joy S. Reidenberg, Ph.D.
Associate Professor
Center for Anatomy and Functional Morphology
Mount Sinai School of Medicine
New York University School of Medicine
New York, New York

William J. Richtsmeier, M.D.
Chief
Otolaryngology Service
Bassett Healthcare
Cooperstown, New York

Hector P. Rodriguez, M.D.
Assistant Professor
Department of Otolaryngology–Head and Neck Surgery
Director of Rhinology
College of Physicians and Surgeons
Columbia University
Presbyterian Hospital
New York, New York

David Rosenstreich, M.D.
Department of Medicine
Albert Einstein School of Medicine
Montefiore Medical Center
Bronx, New York

John J. Rosowski, Ph.D.
Professor
Department of Otology and Laryngology
Harvard Medical School
Massachusetts Eye and Ear Infirmary
Boston, Massachusetts

Jackson Roush, Ph.D.
Professor and Director
Division of Speech and Hearing Sciences
School of Medicine
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina

Robert J. Ruben, M.D.
Distinguished University Professor
Department of Otolaryngology–Head and Neck Surgery
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

John S. Rubin, M.D., F.A.C.S., F.R.C.S.
Consultant Surgeon
The Royal National Throat, Nose and Ear Hospital
Division of The Royal Free N.H.S. Trust
London, United Kingdom

Leonard P. Rybak, M.D., Ph.D.
Distinguished Professor
Department of Surgery
Division of Otolaryngology
Southern Illinois University School of Medicine
Springfield, Ilinois

Ronald C. Scherer, Ph.D.
Professor
Department of Communication Disorders
Bowling Green State University
Bowling Green, Ohio

Gregory J. Schilero, M.D.
Assistant Professor
Department of Medicine
Mount Sinai School of Medicine
Mount Sinai Medical Center
New York, New York

James E. Schwob, M.D, Ph.D.
Professor and Chair
Department of Anatomy and Cellular Biology
Tufts University School of Medicine
Boston, Massachusetts

Nancy Sculerati, M.D.
Pediatric Otolaryngologist (retired)
New York University School of Medicine
New York, New York

Carl E. Silver, M.D.
Professor
Department of Surgery–Head and Neck
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

Adam Silvers, M.D.
Director of Neuroradiology
Next Generation Radiology
Great Neck, New York

Lewis P. Singer, M.D.
Professor
Department of Pediatrics
Albert Einstein College of Medicine
Children's Hospital
Montefiore Medical Center
Bronx, New York

Derek D. Sloan, M.D.
Senior Postdoctoral Fellow
Department of Laboratory Medicine
University of Washington
Seattle, Washington

Richard V. Smith, M.D.
Vice Chairman
Associate Professor
Department of Otolaryngology–Head and Neck Surgery
Medical Arts Pavilion
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

Ruy Soeiro, M.D.
Professor
Department of Medicine
Albert Einstein College of Medicine
Montefiore Medical Center
Bronx, New York

Peter Som, M.D.
Professor
Department of Radiology and Otolaryngology
Director
Head and Neck Radiology
Mount Sinai School of Medicine
Montefiore Medical Center
New York, New York

Bettie Steinberg, Ph.D.
Associate Director
Institute for Medical Research at North Shore
Professor, Department of Ololarnyngology
Long Island Jewish Medical Center
New Hyde Park, New York

Howard D. Stupak, M.D.
Private Practice
Facial Plastic Surgery
New Haven, Connecticut

Mark D. Suski, M.D.
Private Practice
Westlake Village, California

Giovana Thomas, M.D.
Assistant Professor
Department of Otolaryngology–Head and Neck Surgery
Division of Head and Neck Surgery
Miller School of Medicine
University of Miami
Miami, Florida

John S. Touliatos, M.D.
Private Practice
Memphis, Tennessee

Ashok Vaghjimal, M.D.
Private Practice
Allergy, Asthma, and Infectious Disease
Northport, Alabama

Eric P. Voigt, M.D.
Clinical Instructor
Department of Otolaryngology
New York University School of Medicine
NYU Medical Center
New York, New York

Susan B Waltzman, Ph.D.
Professor
Department of Otolaryngology
Director
Cochlear Implant Program
New York University School of Medicine
New York, New York

Ivan Wayne, M.D.
Department of Otorhinolaryngology
University of Oklahoma Health Sciences Center
Oklahoma City, Oklahoma

Jeffrey Wolfe, M.D.
Assistant Professor
Department of Otorhinolaryngology–Head and Neck Surgery
University of Maryland School of Medicine
University of Maryland Medical Center
Baltimore, Maryland

Richard J. Wong, M.D.
Memorial Sloan-Kettering Cancer Center
New York, New York

Peak Woo, M.D.
Professor
Department of Otolaryngology
Mount Sinai School of Medicine
Mount Sinai Medical Center
New York, New York

Barbara A. Zeifer, M.D.
Vice Chairman
Department of Radiology
Beth Israel Medical Center
New York, New York

Karen B. Zur, M.D.
Assistant Professor
Department of Otorhinolaryngology–Head and Neck Surgery
University of Pennsylvania School of Medicine
Children's Hospital of Philadelphia
Philadelphia, Pennsylvania

Part I

THE BASIC PRINCIPLES

Part I
THE BASIC PRINCIPLES

This page intentionally left blank

Chapter 1
SURGICAL HEMOSTASIS

CHRISTOPHER HARTNICK AND HINRICH STAECKER

Surgical hemostasis is described in Schwartz's *Textbook of Surgery* as a "complex process that prevents or terminates blood loss from the intravascular space, provides a fibrin network for tissue repair, and ultimately, removes the fibrin when it is no longer needed." This chapter begins with an overview of the process by way of review and then provides some guidance toward pre-, intra-, and postoperative management of a given patient with a potential bleeding event.

The process of hemostasis begins at the moment of injury to the endothelial lining of the vascular bed. Left undisturbed, the endothelial cells themselves act within a complex series of events to prevent clotting. When the endothelial wall is disrupted, the various elements of the blood are exposed to the underlying collagen. After this initial event, several events rapidly occur to stem the flow of blood from the wound. The first process is that of vasoconstriction at the level of the capillary bed. This process is dependent upon the local contraction of smooth muscle and is influenced by the next event in the cascade of hemostasis; namely, platelet aggregation, as thromboxane A2, a powerful vasoconstrictor, is produced by the release of arachidonic acid from platelet membranes during aggregation.

Within 15 seconds after the onset of vasoconstriction, platelets begin to aggregate as they stick to the exposed collagen. The platelets adhere to the wound bed and begin to form a plug, which is the initial matrix upon which fibrin will eventually deposit. As the platelets are beginning to aggregate and adhere to the subendothelial collagen, the intrinsic and extrinsic pathways of the coagulation system are also activated by damage to the endothelium, and the two cascades move toward the end point where prothrombin is converted to thrombin, which in turn catalyzes the conversion of fibrinogen to fibrin. Insoluble fibrin is deposited in and around the platelet plug, and a more fully developed clot is formed. At any point on the path from endothelial injury to thrombus formation, a host of factors can derail the process of hemostasis and can produce a potential for prolonged bleeding.

CLINICAL EVALUATION OF A PATIENT FOR POTENTIAL HEMOSTATIC DISORDER

All patients who are scheduled for surgery or who present with an episode of bleeding should be evaluated for a potential occult bleeding disorder. The first and perhaps most sensitive screen to identify a bleeding disorder is the taking of a careful history. A pattern of easy bruising, of prolonged bleeding after minor or major surgery or after tongue biting, of heavy menstrual bleeding, or of any family history of excessive bleeding all warrant further pursuit of an underlying

TABLE 1–1 CLOTTING DISORDERS

Disorder	Factor	Testing	Treatment	Inheritance	Comments
Hemophilia A	VIII	PTT, factor VIII	Cryoprecipitate	Sex-linked recessive	
Hemophilia B	IX	PTT	FFP	Sex-linked recessive	
von Willebrand's disease	vWF	Bleeding time, PTT	DDAVP, cryoprecipitate	Autosomal dominant	Variable presentation
Factor XI deficiency	XI	PTT	FFP		
Factor XII deficiency	XII	PTT	FFP		No tendency to bleed

DDAVP, Desmopressin; FFP, fresh frozen plasma; PTT, partial thromboplastin time; vWF, von Willebrand factor.

problem. A history of bleeding disorders such as von Willebrand's disease and hemophilia should be noted. Any chronic medical problems such as liver or renal disease should be noted, as should any medications that might affect hemostasis (see **Tables 1–1** and **1–2**).

Once a careful history has been taken, the physical exam should be tailored to identify any hematologic abnormalities. Small telangiectatic lesions on the face, oral or nasal mucous membranes, or fingertips are suggestive of hereditary hemorrhagic telangiectasia; perifollicular skin hemorrhage suggests scurvy, and hemarthrosis in the absence of trauma suggests hemophilia.

LABORATORY EVALUATION

After a careful history and physical examination, the question arises as to what blood tests or further studies are required to fully assess the patient's relative risk of bleeding. To some extent, the decision hinges upon both the patient and the extent of surgery that is planned. Patients who are actively bleeding at the time of presentation or who are being scheduled for major surgery routinely require a complete blood count (CBC), a prothrombin time/partial thromboplastin time (PT/PTT), and a full series of blood chemistries, including liver function studies. Further workup based on the history and physical or laboratory abnormalities may merit a hematology consultation for guidance. For patients undergoing more minor procedures, the workup can be tailored by the relative risks of the patient and the surgery itself, although what constitutes "relative" remains broadly interpreted. A good case in point is the debate over what laboratory values are needed prior to performing a tonsillectomy. Although

TABLE 1–2 ACQUIRED DISORDERS OF COAGULATION/TESTING

Heparin	Glycosaminoglycan that binds antithrombin III, resulting in inhibition of thrombin. Prolongs PTT, can be reversed with protamine
Coumadin	Impairs synthesis of vitamin K–dependent factors, prolongs PT
Streptokinase	A bacterial protein that enhances activation of plasmin, resulting in lysis of fibrin
Urokinase	Directly cleaves plasminogen to form plasmin
Tissue plasminogen activator	Less prolongation of PTT but has an increased risk of intracranial bleed
Aspirin	Irreversibly inhibits production of thromboxane A2
Renal failure	Platelet and small vessel dysfunction: treat with DDAVP
Massive blood transfusion	Bleeding probably due to inadequate platelet function; acidosis and hyperthermia may aggravate the situation
Disseminated intravascular coagulation	Consumptive coagulopathy initiated shock due to infection or a variety of other causes; PT/PTT prolonged, elevated fibrin degradation products. Treat underlying cause, replace blood, give cryoprecipitate and FFP as needed
Medications	NSAIDs (e.g., Toradol), cephalosporins, dextran

DDAVP, Desmopressin; FFP, fresh frozen plasma; NSAIDs, nonsteroidal anti-inflammatory drugs; PT, prothrombin time; PTT, partial thromboplastin time.

the American Academy of Otolaryngology–Head and Neck Surgery currently recommends that coagulation studies are warranted only in patients with positive histories or physical examinations, many otolaryngologists do not hold this "standard of care," even in the face of well-conceived prospective studies.

The following descriptions outline the commonly ordered tests and highlight their clinical importance.

Tests for Platelet Evaluation

1. Platelet count
2. Peripheral blood smear
3. Bleeding time

Platelets are 2 μm fragments of megakaryocytes that normally number 200,000 to 400,000/mm^3. The life span of a platelet ranges from 7 to 9 days. A routine platelet count will give some indication as to the number of circulating platelets. If the platelets are recorded as "clumped," or if there is some question as to the accuracy of the count, a peripheral blood smear can be performed, and the platelets can be manually counted. In cases where bleeding disorders, such as von Willebrand's disease, are suspected, a "bleeding time" can provide useful information as to the ability of a patient's blood to form a clot. Using the Ivy technique, a normal bleeding time averages 5 $\pm$ 2 minutes.

Test of Coagulation Pathway

The PT/PTT and international normalized ratio (INR) are designed to test the intrinsic and extrinsic cascades, which are part of the coagulation pathway. The PT tests the factors involved in the extrinsic pathway; namely, factors II, VII, IX, and X, which are produced by the liver. The PTT tests the factors in the intrinsic pathway. The INR was introduced because of laboratory variability in reporting the PT. The INR incorporates a correction factor into the PT ratio and standardizes the results. Specific tests for levels of each one of the clotting factors are available and useful in particular cases.

COMMONLY SEEN BLEEDING ABNORMALITIES

Platelet Disorders

Thrombocytopenia is the most common hematologic cause of perioperative bleeding. Thrombocytopenia can arise secondarily to occult disease, megaloblastic anemia (B_{12} folic acid deficiency), from uremia, from certain drugs, or from massive blood loss requiring transfusions. Exchange of one blood volume (11 units for a 75 kg male) will result in a decrease in platelet count from 250,000 to 80,000/mm^3. Loss of platelets can also be caused by drug allergies or diseases such as idiopathic thrombocytopenic purpura (ITP). As long as the platelet count is $>$ 50,000/mm^3, there is no absolute need for transfusion. Once the platelet count drops below 40,000/mm^3, the risk of spontaneous bleeding increases. The treatment of thrombocytopenia in the nonacute setting begins with an attempt to identify and remedy the causative factor. If the cause is either alcohol or viral related, then the platelet count should return to normal 1 to 3 weeks after the inciting factor has been removed. In the acute setting, where platelets are needed emergently, platelets can be transfused. One unit of pooled platelets usually raises the platelet count by 10,000; therefore, 6 to 8 units are usually required to restore normal clotting.

There are other platelet disorders that do not manifest as thrombocytopenia, but rather are functional disorders (suggested by normal platelet count and increased bleeding time). The most common of these disorders is related to aspirin usage. Aspirin inhibits the entire prostaglandin pathway by irreversibly acetylating cyclooxygenase, which is involved in platelet aggregation. The process is irreversible, so the circulating platelets must be replenished (in a process that takes roughly 72 hours) before they can again function normally. Cephalosporins have also been suggested to cause platelet dysfunction and should be considered a potential cause of bleeding disorders if no other causes are found.

Another common platelet disorder is seen in von Willebrand's disease; the von Willebrand factor (vWF) normally allows platelets to adhere to the subendothelial system and is responsible for carrying the coagulant portion of factor VIII. When vWF is missing or defective, the ability of platelets to form a plug and begin the process of hemostasis is curtailed. The diagnosis is suggested by a prolonged bleeding time, and treatment may require cryoprecipitate or desmopressin (DDAVP), which causes a transient release of vWF from endothelial cells. This disorder is usually first noted in childhood, and presentation can be variable, depending on the amount of functional vWF. For surgical treatment, vWF should be maintained at 50% of normal.

DISORDERS OF THE COAGULATION SYSTEM

As with platelets, there are a host of factors that influence and can alter the coagulation system. These include chronic diseases (notably liver disease because factors II, VII,

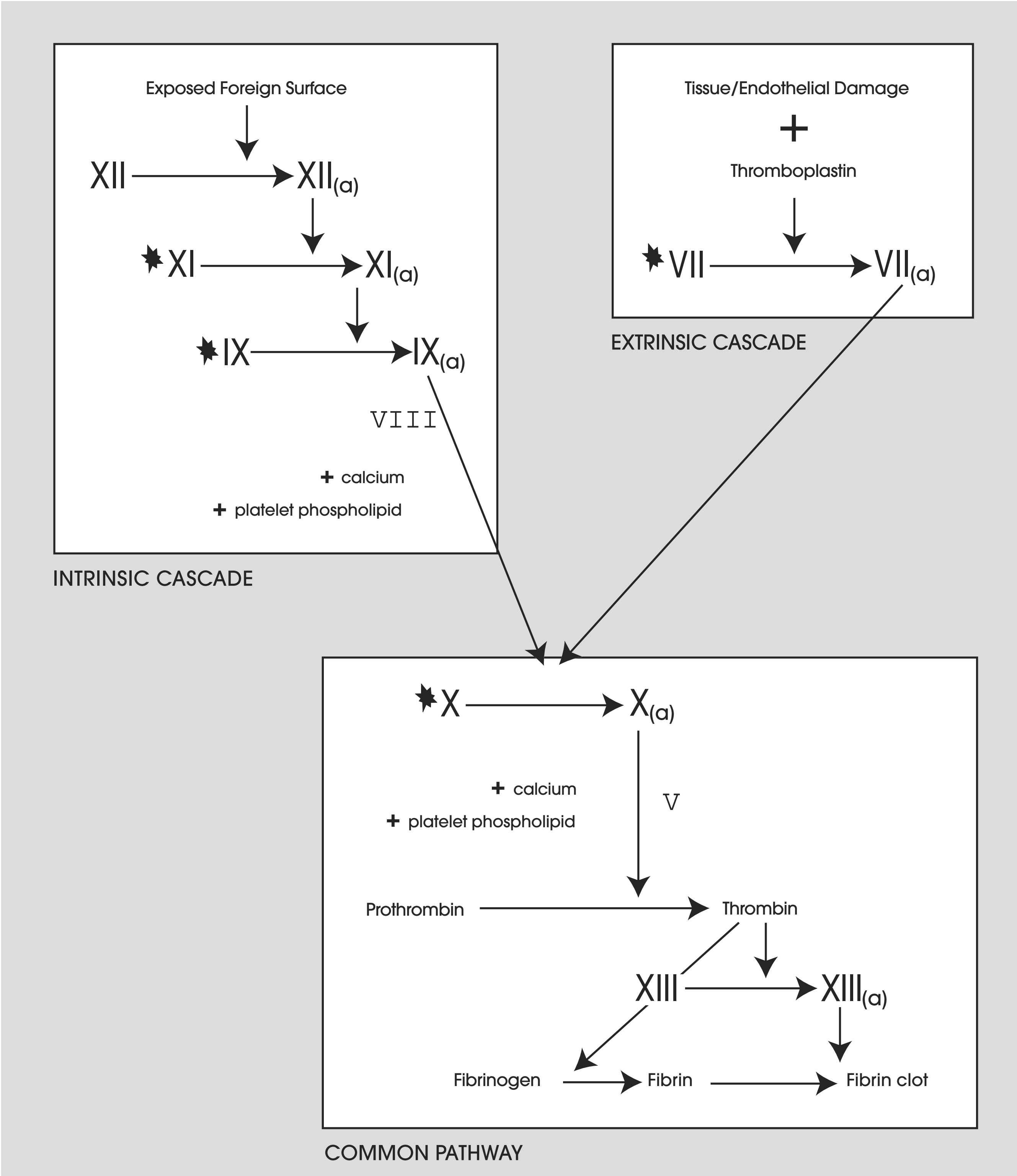


Figure 1–1 *Overview of intrinsic, extrinsic, and common pathways of hemostasis. Vitamin K–dependent factors are marked with a star (serine proteases). The intrinsic pathway involves the sequential activation of factors XII, XI, and IX (Christmas factor), leading to activation of factor VIII, leading to the activation of the X/V complex of the common pathway (this pathway is tested with the PTT). This leads to activation of the common pathway in which the X/V complex converts prothrombin (factor II). Thrombin functions to convert fibrinogen to fibrin (factor I), as well as to activate factor XIII, which polymerizes fibrin to form a clot. Thrombin also indirectly activates protein C, which inhibits the X/V complex.*

IX, and X are made in the liver), drugs, and massive blood loss followed by transfusion. One of the most common diseases of the coagulation system is an inherited deficiency of one or more of the factors involved in either the intrinsic or extrinsic pathways. Examples of these disorders are hemophilia (deficient for factor VIII) and Christmas disease (deficient for factor IX). Treatment of these disorders acutely may require the administration of fresh frozen plasma.

Another common disorder of the coagulation system is iatrogenic manipulation for purposes of anticoagulation (**Fig. 1–1**).

The issue of aspirin has been discussed previously because it affects platelet aggregation. Other common medications used to affect the hemostatic system are heparin and warfarin sodium (Coumadin). Heparin is a mucopolysaccharide extracted from the mast cells. It exerts its actions in several ways: by slowing the conversion of prothrombin, by potentiating the effect of antithrombin III, and by decreasing the degree of platelet adhesiveness. It is administered intravenously and has a half-life of 90 minutes. It is monitored by following the PTT level. Due to the short half-life of heparin, discontinuing it several hours before a given procedure should allow the PTT to normalize. More rapid equilibration may require the administration of protamine sulfate.

Warfarin sodium is the oral substitute for heparin. It functions by inhibiting the production of the vitamin K–dependent factors of the coagulation cascade (namely, II, VII, IX, and X). Its half-life is 36 hours. The effects of warfarin sodium can be monitored by following the PT level and the INR. Normalization of the PT and INR can be affected by the administration of fresh frozen plasma.

EVALUATION AND MANAGEMENT OF PERIOPERATIVE BLEEDING

The evaluation of a patient with a bleeding episode, be it on admission to an emergency room or after an operative procedure, depends in its depth on the acuity and severity of the episode. As in all patient management, there needs to be a primary and a secondary survey. The primary survey consists of evaluating the "ABCs" (the airway, breathing, and circulation) and managing the patient accordingly. Once the patient has been stabilized, a more thorough review of the patient can be accomplished. Factors involved in excessive bleeding include ineffective local hemostasis, complications of blood transfusions, hematologic abnormalities, and consumptive coagulopathies.

Control of localized bleeding begins with pressure applied to the area. If there is an identifiable vessel, it can be ligated or cauterized, or the area can be packed. All packing transmits pressure to the wound bed and provides a scaffold to augment the hemostatic process. Within the realm of otolaryngology, various chemical packing agents are commonly used: these include Gelfoam, Oxycel, Surgicel, and Avitene, among other products. Gelfoam is made from denatured animal skin gelatin. It acts as a pressure matrix; when combined with topical thrombin, a hemostatic effect is produced. Oxycel and Surgicel are cellulose materials that produce a hemostatic effect by their interaction with blood products to form a "clot." Avitene is microcrystalline collagen that can be helpful with a diffusely oozing wound bed.

When local control cannot be obtained, the decision must be made whether the amount of bleeding warrants surgical exploration and control or whether the difficulty stems from an underlying hematologic problem that needs to be addressed (**Fig. 1–2**). The laboratory values can be helpful in this regard because they can guide management. If the patient has received massive transfusions, this also must be kept in mind because the patient may require additional factors (i.e., platelets and fresh frozen plasma). If the patient is septic, a consumptive coagulopathy may develop and needs to be treated accordingly by attempting to treat the source of the infection as well as by replacing the various hemostatic factors.

Overall, evaluation and treatment of a patient with either a potential or an active bleeding disorder require some knowledge of the hemostatic process by which the human body repairs itself. Such knowledge allows the development of an algorithim to assess each aspect of this process. Initial management will either quell or temporize the problem; more complicated problems

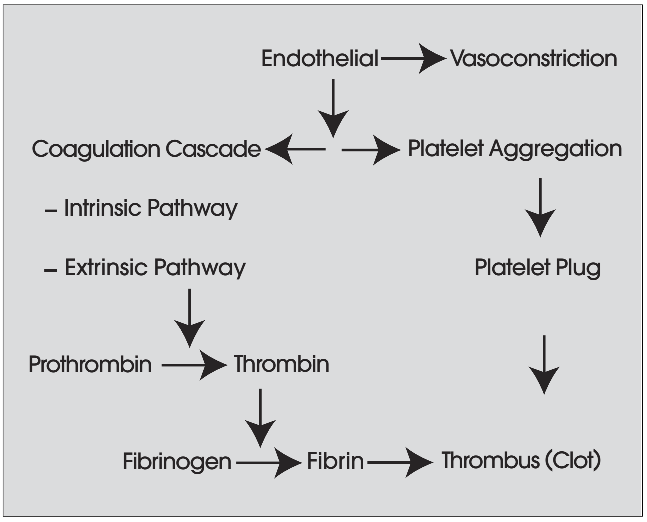


Figure 1–2 *Schematic review of initiation and progression of hemostasis.*

merit surgical intervention and the treatment of the underlying hematologic abnormalities.

Suggested Readings

Bentler E, Lichtman MA, Coller B, Kipps T. Williams Hematology. New York: McGraw-Hill; 1995

Close HL, Kryzer TC, Nowlin JH, Alving BM. Hemostatic assessment of patients before tonsillectomy: a prospective study. Otolaryngol Head Neck Surg 1994;111(6):733–738

Cohen JR. Vascular Surgery for the House Officer. 2nd ed. Baltimore: Williams and Wilkins; 1992

Schwartz S, ed. Principles of Surgery. New York: McGraw-Hill; 1991

Self-Test Questions

For each question select the correct answer for the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. The most common cause of a platelet-related clotting disorder is
   - A. Idiopathic thrombocytopenic purpura
   - B. Megaloblastic anemia
   - C. Von Willebrand's disease
   - D. Aspirin therapy
2. Coumadin inhibits the production of which vitamin K–dependent clotting factors?
   - A. II, V, XIII
   - B. IV, VII, X
   - C. II, VII, IX, X
   - D. II, VII, VIII
3. The best treatment for severe von Willebrand's disease is
   - A. Platelet transfusion
   - B. Protamine infusion
   - C. Factor VIII infusion
   - D. Cryoprecipitate

Chapter 2

WOUND HEALING

JANE A. PETRO, MARK D. SUSKI, AND HOWARD D. STUPAK

Many animals, like lizards and stone crabs, heal tissue injury through a process of regeneration. Regeneration occurs in mammalian fetal healing, but mature tissues in humans respond to injury through the formation of scar tissue. Scar tissue serves to restore integrity and function of injured tissue, but the healed tissues may be impaired in shape and appearance, and may even be limited in function because of this scar tissue. Thus scar tissue formation does not represent the ideal outcome of wound healing. Mammalian tissue healing is principally described as having four stages: hemostasis, inflammation, cellular proliferation, and maturation. These occur in an orderly fashion, beginning at the time of injury (or illness) and proceeding through successive steps to the production of a mature, stable scar. The ideal goal of surgical healing would be to achieve timely regeneration, but such wound healing has not been achieved in the adult as yet. This chapter will review the processes involved in normal healing of skin, bone, cartilage, and mucosa. Tissue engineering, the creation of replacement tissues, and fetal healing will also be reviewed, albeit briefly.

HISTORY AND PROGRESS OF WOUND HEALING

A review of our understanding of wounds and healing extends back to the earliest documented medical writings. Progress in wound healing, surgical techniques, and management of trauma is frequently linked to war because the volume of patients encountered contributes to observation of injury and to the management of frequently seen wounds. In association with that, wounds have long been a fascination of both poetic and medical authors. Egyptian papyri, the Bible, and Homer's *Iliad* all contain descriptions of various wounds and suggested remedies. Seven of the 48 cases discussed in the Smith papyrus involve discussion of wounds and their management. In Homer's *Iliad* and *Odyssey*, the wounds described included dislocations, treatable by closed reduction, and sword and spear injuries treated by removal of the penetrating weapon, as well as simple dressing of the wounds. Of the 147 wounds mentioned in the *Iliad*, the mortality rate was 77.6%. Hippocrates described the reduction of fractures and treatment of arrow and sword injuries, and emphasized the importance of making a prognosis for the patient based on the nature of the wound. Basic principles of suturing, removal of foreign bodies, drainage of abscesses, and the sacred nature of the relationship between the physician/ surgeon and the patient were part of the medical canon. For a detailed and very entertaining history of wound management, interested readers should refer to Guido Manjo's book, *The Healing Hand: Man and Wound in the Ancient World* (1975).

Little progress in wound management occurred between the time of Galen (c. AD 130–200), a Greek physician whose writings on medicine became the standard canon of medicine), and the 15th century. Galen reported in an encyclopedic fashion on the medical knowledge of his day, basing his work on his own observations as an anatomist and vivisectionist. He had access to thousands of animals and the gladiators of the Roman games, but he derived his principles from the Hippocratic theories of the four humors. Despite his role in the active treatment of many wounds, he offered no effective remedies or treatments. The use of bread and wine as dressings, as described by Greek authors, at least caused less injury than other prescriptions of the times. Galen, relying on Hippocratic dicta, emphasized the value of starving, bleeding, and purging. This information was accepted and used uncritically for the next 1000 years. Galen's work was finally supplanted by the discovery of the Greek author Aulus Cornelius Celsus (c. 25 BC–AD 50). Celsus was also an encyclopedist, working ~100 to 150 years before Galen. Because he was writing in Greek rather than Latin, his work rested in obscurity until a copy, 500 years old at the time, was discovered in the Basilica of St. Ambrose in Milan. Celsus's encyclopedia included eight books on medicine. *De medicina* became one of the earliest medical books to be printed (1478) after the invention of moveable type. Of interest to surgeons and students of wound healing is the classic description Celsus left us of inflammation: "Now the characteristics of inflammation are four: redness and swelling, with heat and pain."

These initial clinical observations regarding healing in the earliest medical writing were then described in increasing detail and with improved results. Surgeon-scientists like Ambrose Pare, John Hunter, Joseph Lister, and Alexis Carrel advanced wound management, increasing the survival from wounds, and permitting rational, science-based methods of wound care. From the macro-descriptions and observationally prescribed remedies of earlier times, the modern introduction of a scientific method, coupled with tools like the microscope, permitted observation of the cellular events involved in wound healing. This, in turn, permitted the specific description of wound healing still in use today. While studying the relationship of healing and tensile strength, Howes, Sooy, and Harvey, in 1929, identified the three stages of wound healing, which they labeled inflammation, fibroplasia, and maturation. Clinical observations of the gross events (inflammation, cicatrix) related to healing were first supplemented by microscopic descriptions of cells migrating into the site of injury. These observations have since progressed to the elucidation of the complex biochemical cellular processes of growth factor production, cellular and extracellular matrix interactions, and the genetic codes for differentiation and maturation, with their promise of tissue repair by coordinated regeneration, as is seen during early fetal development.

Tissue injury is the real first stage of wound healing. Clean, sharp cutting injuries result in a healing process which is remarkably different from that resulting from contaminated, crushing, or multisystem tissue trauma. The results of healing from these different wounds are themselves different and predictable. In the first phase of healing (hemostasis), injury is accompanied by bleeding, which results in formation of a blood clot. This is followed by a variable period of inflammation lasting for 1 to 3 days, in the absence of infection, or it may last as long as infection is present. Pre-Listerian wound observations distinguished between "laudable" pus and other pus that was invariably associated with mortality. The creamy, thick pus considered laudable was usually primary staphylococcal infection accompanying non-antiseptic skin injury. Such infections might lead to further complications, but they usually healed eventually. Thin, watery, malodorous pus (associated with streptococcal or gram-negative infections) predicted mortality and remains today a significant contributor to the morbidity and mortality associated with trauma and hospitalization.

The introduction of gunpowder into warfare resulted in more complex and severe injuries. The inevitable association of infection with gunshot injury, and the frequency of death, resulted in the "heroic" surgery of the battlefield encountered in Ambrose Pare's time. The use of boiling oil and cautery with hot coals or heated instruments added burn to blast. After running out of the usual recommended remedies, Pare, in *The Treatment of Wounds Caused by Firearms* (1547), noted that the "mistreated" soldiers actually did better than those whose wounds were boiled and packed with (very dirty) cotton. His observation that "man may dress the wound, but only God can heal it" implies the principle that less intervention is better than more. This aphorism applied until Lister introduced the practice of wound cleansing. Prior to Lister, compound fractures were routinely treated with amputation. His series of 13 patients whose fractures were first packed with carbolic acid, then reduced, led to one amputation and 12 successfully healed fractures and wounds. This remarkable discovery, in association with the introduction of pain control with anesthesia, made possible both modern elective surgery and the possibility of survival after many once mortal injuries.

No discussion of the evolution of surgery, especially that related to the head and neck, can be complete without mention of the work of Gaspard Tagliacozzi. He described the "Italian method" of nasal reconstruction in *De Curtorum Chirurgia* (1597). Using skin from the upper arm and a delayed attachment method, his operation foreshadowed the reconstructive techniques perfected during World War I by Sir Harold Gilles and his colleagues at Queens Hospital at Sidcup, England. Gilles and his colleagues, an assortment of dentists, surgeons, and anesthesiologists, coming from both the United States and England, developed the specialty of reconstructive surgery out of necessity. They created an approach dedicated to treating the complex facial injuries associated with aircraft and motorized vehicle crashes and burns, in addition to those injuries caused by guns and explosions, as seen in previous warfare. Gilles's principles of practice based primarily on pedicled flap reconstruction emphasized attention to detail with a multidisciplinary, team-oriented approach, which led to successful outcomes in thousands of cases of war-injured veterans.

Alexis Carrel, winner of the Nobel Prize in medicine in 1912, wrote *The Treatment of Infected Wounds* (1917). Carrel emphasized aseptic technique and developed microvascular methods still used today for repair of blood vessels. His experimental studies into surgical procedures on the heart and great vessels led him into early work on transplantation, techniques doomed to failure prior to an improved understanding of the immune system. But his work laid the foundation for vascular and cardiac surgery today. His early work on wound infection emphasized the value of delayed secondary closure, another technique still applicable today.

Modern wound care combines the control of the complex molecular events involved in healing through pharmacological and nutritional manipulation, with improved surgical technique. These advances have resulted in more rapid closure of difficult wounds and improved aesthetic results in simpler ones.

CELLULAR BASIS OF WOUND HEALING

Contemporary understanding of wound healing has identified the stages of healing, with the cellular events that accompany each stage. These stages can be divided into several classifications. Because wounding plays such a crucial role in the outcome of healing, wounding or bleeding is often included as one of the stages. For the purposes of this discussion, wound healing will be divided into four phases: hemostasis, clot formation after wounding; inflammation, the cellular events that follow wounding; proliferation, the production of collagen and the extracellular matrix that establishes the scar; and maturation, the establishment of a balance between the scar and the healing process. The classic graph of the events is seen in **Fig. 2–1**. The changes in cell type and their relative concentration in the wound following injury are depicted relative to the stages of wound healing and the days postinjury. These curves will vary in the presence of delayed wound healing, or infection, which prolongs inflammation, and with wound defect characteristics and closure techniques.

Each phase of wound healing is characterized by a distinct cellular appearance, with unique cells and

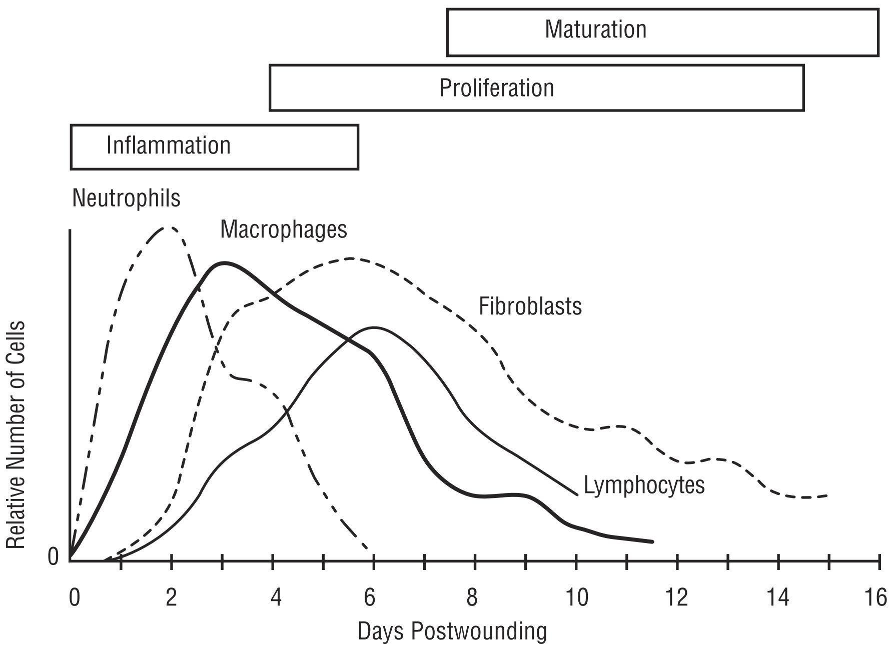


Figure 2–1 *Changes in cell types at a wound site following injury are relative to the stage of wound healing and time postwounding. These curves vary in the presence of delayed wound healing and infection, both of which prolong the inflammation phase of healing.*

biochemical markers. **Table 2–1** correlates the phase of healing with those cells and a few of their products.

Hemostasis

The bleeding that accompanies injury initiates the inflammatory process that is a prerequisite to successful healing. The mechanism by which bleeding is controlled begins with (1) vasoconstriction (an effect of norepinephrine and epinephrine) that slows the flow of blood; (2) platelet plugging, which acts to cork the flow; and (3) fibrin clot formation, which also traps red cells, enhancing the cork effect. These serve to mechanically control bleeding and provide a lattice-like framework to support the migration of subsequent cellular infiltrates. The intrinsic coagulation path is activated via factor XII, when blood is exposed to foreign surfaces. This branch of the process is not essential. The alternative, the extrinsic pathway, is initiated by tissue factor, binding to factors VII and VIIa. Factor XIII (fibrin-stabilizing factor) initiates fibrin clot development. This step is crucial to the initiation of wound healing. These hemostatic factors are found on extravascular cellular surfaces. Platelets also activate the intrinsic coagulation cascade by their response to exposure to subendothelial collagen, which in turn stimulates platelet aggregation. Platelet α granules release cytokines and growth factors such as serotonin, fibronectin, platelet-derived growth factor (PDGF), transforming growth factor α (TGF-α), and platelet activating factor (PAF). Vasoconstriction of capillaries reverses 15 minutes after injury, mediated by histamine, kinins, prostaglandin, and leukotrienes. The subsequent increased flow, associated

Table 2–1 The phases of wound healing, with their distinctive cellular and biochemical components

Phase	Cells	Biochemical Hallmarks
Hemostasis	Platelets	Fibrin, cytokines, growth factors
Inflammation	Neutrophils	Fibrinogen, proteases, interleukins, growth factors
	Macrophages	Cytokines, growth factors
	Lymphocytes	Cytokines
	Mast cells	Unknown
Proliferation	Fibroblasts	Proteoglycans, collagen deposition, fibronectin
	Epithelium	Keratin, growth factors
	Endothelium	Growth factors
Remodeling		Collagen fibril cross-linking

with gaps in the endothelium of the capillaries, facilitates the migration of neutrophils into the wound area and releases plasma from the intravascular space, providing albumin and globulin, which combine with fibronectin and fibrin in the formation of the provisional matrix. Complements C3a and C5a also increase capillary permeability and are chemotactic to neutrophils and monocytes.

Both the intrinsic and extrinsic coagulation pathways stimulate formation of thrombin, which then serves as a catalyst to the conversion of fibrinogen to fibrin. The thrombin clot within the wound provides the provisional matrix for the extracellular matrix (ECM) that becomes the scaffolding for subsequent cellular migration, adhesion, and proliferation, critical activities in the healing process. Impaired wound healing occurs in the absence of clot formation, such as factor XIII (fibrin-stabilizing factor) deficiency, by several critical steps. Such impairments include decreased chemotaxis and diminished cell ability to adhere to the fibrin matrix, either of which will result in reduced cell migration. The identification and elucidation of the role played by growth factors have been the most recent major advances in the knowledge of wound healing. Growth factors may best be described as tissue-specific polypeptides that act as local regulators of cellular activity. Growth factors exert their biological function by binding specifically to large cell surface transmembrane receptors on their target cells. **Table 2–2** identifies many of the known growth factors of this initial phase of wound healing and their activities during this time.

Growth factors are identified by their cell source and their function, but these are multifactorial, depending on the extracellular milieu, concentration of other factors, and density of cells in the vicinity. **Table 2–3** lists some of the more thoroughly studied growth factors and identifies their cell sources, as well as their activities that have been identified to date.

The activity of these cells and their secretory products are mediated by cell surface receptors, which regulate stimulus and response selectively. Many of these factors have more than one function and effect and can influence different cells over time and at different concentrations. Chemoattraction of cells into the extracellular matrix formed by the fibrin scaffold is followed by the activation of those cells. Thus a carefully regulated progression of cells, their changing function, and their differentiation serves to orchestrate the wound healing process, and hemostasis prepares the way for the next stage of healing, inflammation.

The Inflammatory Phase

The inflammatory phase begins immediately upon injury and lasts for between 2 and 5 days, unless infection develops, in which case it can be prolonged indefinitely. If clotting proceeds normally, the first cells migrating into the wound appear almost simultaneously to initiate the inflammatory process. Neutrophils follow platelets into the wound area, guided by chemoattractant factors including complement, interleukin-1, tumor necrosis factor α (TNF-α), TGF-β, and platelet factor 4. Neutrophils debride the wound by phagocytosis of foreign matter, damaged cells, and bacteria. Though

Table 2–2 Factors involved in initiating wound healing

Hemostatic Factors	**Function**
Histamine	Capillary vasodilatation and increased permeability
Plasma fibronectin, fibrin	Adhesion, chemoattraction, coagulation, scaffolding for cell migration
Factor XIII	Induces adhesion and chemoattraction
Circulatory growth factors	Regulate chemoattraction, fibroplasia, mitogenesis
Complement	Stimulates antimicrobial activity, chemoattraction
Platelet-Derived Factors	**Function**
Cytokines and growth factors	Regulation of chemoattraction, fibroplasia, and mitogenesis, ligand for platelet aggregation and matrix formation
Fibronectin	Platelet aggregation
Platelet-activation factor	Chemotaxis, platelet aggregation, vasoconstriction
Thromboxane A2	Chemotactic attraction for fibroblasts, monocytes
Platelet factor IV	Neutralize heparin activity, inhibit collagenase; chemotactic for neutrophils, induce vascular permeability
Serotonin	Stimulate cell proliferation and migration, induce platelet aggregation
Adenosine dinucleotide	Stimulate cell proliferation and migration, induce platelet aggregation

TABLE 2–3 GROWTH FACTOR ACTIVITY IN WOUND HEALING

Growth Factors	Cell Source	Activity
TGF-α	Platelets, macrophages, keratinocytes	Activates neutrophils, mitogen for fibroblast stimulates angiogenesis
TGF-β	Platelets, macrophages, lymphocytes	Stimulates fibroplasia and angiogenesis, induces proliferation of many different cells
PDGF	Platelets, macrophages, keratinocytes, endothelial cells	Chemoattractant for neutrophils, fibroblasts, mitogen for smooth muscle cells and fibroblasts
FGF	Macrophages, neural tissue, nearly ubiquitous	Stimulates endothelial cell growth, mitogen for mesodermal and neuroectodermal-derived cells
EGF	Platelets, keratinocytes, salivary gland	Mitogen for keratinocytes, endothelial cells, and fibroblasts
IGF	Liver	Mitogen for fibroblasts; stimulates smooth muscle cells, lymphocytes, and chondrocytes

EGF, epidermal growth factor; FGF, fibroblast growth factor; IGF, insulin-like growth factor; PDGF, platelet derived growth factor; TGF, transforming growth factor.

they are the predominant cell for the first 48 hours, they are not essential for uncomplicated wound healing. Neutrophils also produce proinflammatory cytokines, which are chemoattractant to fibroblasts and keratinocytes. In the presence of infection, bacterial products may also stimulate neutrophil migration.

Hunt (1976) characterized the role of the macrophage in this process as the conductor of wound healing. The primary functions of macrophages that have been demonstrated to date include phagocytosis and wound debridement, cellular recruitment and activation, angiogenesis, and the regulation of matrix synthesis. Macrophages are both the phagocytic cells responsible for wound debridement and the secretory source of the cytokines that regulate angiogenesis and fibroplasia. They are also known to play a critical role in the production of nitric oxide, an antimicrobial. In the absence of nitric oxide, several animal models have demonstrated significant wound healing impairment. Experiments using antimacrophage serum document impairment of phagocytosis, retardation of fibroplasia, and decreases in fibrogenesis, which impede wound healing and decrease the rate of gain of tensile strength. The effect of systemic steroids on wound healing similarly inhibits macrophage proliferation, and through them, wound healing. **Table 2–4** lists some of the effector mechanisms of macrophage

TABLE 2–4 MULTIPLE ROLES OF MACROPHAGES

Task	Effectors	Agents
Phagocytosis and antimicrobial function	Oxygen radicals	H_2O_2, O_2, OH, nitric oxide
Wound debridement	Phagocytosis	
	Enzymes	Collagenase, elastase
Angiogenesis	Growth factors	TGF-β, FGF, PDGF
	Cytokines	TNF-α
Matrix synthesis and regulation	Growth factors	TGF-β, EGF, PDGF
	Cytokines	TNF-α, IL-1, IFN-Y
	Enzymes	Collagenase, arginase
	Prostaglandin	PGE2
Cell recruitment and activation	Growth factors	PDGF, TGF-β, EGF, IGF
	Cytokines	TNF-α, IL-1, IL-6
	Fibronectin	

EGF, epidermal growth factor; FGF, fibroblast growth factor; IFN, interferon; IGF, insulin-like growth factor; IL, interleukin; PDGF, platelet derived growth factor; PGE, prostogland-E; TGF, transforming growth factor; TNF, tumor necrosis factor.

function, as well as the cell products that regulate the inflammatory process of wound healing.

The wound matrix changes during this stage. Fibrin and thrombin are influenced by the transudation of plasma into the wound site with deposition of fibronectin and hyaluronic acid and sulfated glycosaminoglycans (GAGs). This matrix stimulates fibroblasts from adjacent tissues to migrate into the field and stimulates fibroblast production of type I collagen, elastin, and proteoglycans. These functions are promoted by TGF-β, initiating the next stage of healing, proliferation.

The Proliferative Phase

This stage of healing lasts for between 2 days and several weeks, depending on inflammation, the action of growth factors, and the rate of epithelialization. Effective primary healing, with edge-to-edge approximation of the wound margins in the absence of infection, would result in the shortest period of proliferation. Prolonged inflammation, associated with infection, an open wound, or continuous disruption, results in alterations of cell sequences in the wound, increased levels of fibrogenic cytokines, and the initiation of fibroproliferative disorders characterized by hypertrophic scars and keloids. The active cell of this phase, the fibroblast, differentiates into a producer of collagen and other extracellular substances, or a smooth muscle cell, the myofibroblast. The fibroblast is responsible for production of GAGs, hyaluronic acid, chondroitin 4-sulfate, dermatan sulfate, and heparin sulfate, substrates for ground substance. In combination with the extracellular deposition of collagen replacing fibrin and thrombin, the ground substance establishes the matrix into which new blood vessels migrate and across which epithelium migrates. Failure of epithelialization leads to increases in inflammatory cells and neoangiogenesis, resulting in the formation of granulation tissue (also called "proud flesh" in the vernacular). Hypoxia, acidosis, and the presence of lactate all stimulate neoangiogenesis.

Fibroblasts, smooth muscle cells, epithelial cells, and endothelial cells all contribute to collagen production. But the role of fibroblasts is the most essential, depending on which tissues are involved in healing. Collagen is the chief structural protein in all connective tissue and provides the tensile strength in skin and bone, as well as the integrity of healed wounds of vessels, nerves, and gut. Intracellular collagen production is a protein synthesis of repeating amino acid groups, Gly-X-Y. The amino acids represented by *X* and *Y* are often lysine and proline. Peptide production proceeds by the formation of α chains. Three distinct polypeptide chains are synthesized in right-hand helical configurations, which are then twisted into a left-handed superhelix. Critical components of collagen synthesis include the hydroxylation of lysine and proline for covalent cross-link formation. Proline is the dominant amino acid, often used as a marker of the quantity of collagen in tissues. Hydroxylation requires cofactors in addition to specific enzymes: Oxygen, vitamin C, ferrous iron, and α-ketoglutarate deficiencies result in underhydroxylation and a weakening of collagen cross-linking, reducing both tensile strength and bursting strength of healing wounds. Steroids inhibit enzyme activity in the wound, with similar results. Once the intracellular production and cross-linking of collagen is completed, the molecules are secreted as procollagen into the extracellular space. Procollagen contains nonhelical extensions of the α chains, which require cleavage by registration enzymes. After cleavage of these ends, successful aggregation of the molecules into fibrils is initiated by proteoglycans in the extracellular matrix. The fibrils link into larger fibers, with the strongest cross-linkages occurring between hydroxylysine residues. The strongest collagen fibers are in those tissues with the highest density of hydroxylysine.

Collagen synthesis and degradation proceed simultaneously, with breakdown of collagen increased during inflammation. Collagenase-specific enzyme activity is highly regulated by cytokines as part of their role in matrix formation. Integrins, cell surface receptors, mediate multiple features of this stage of wound healing, including collagen-dependent adhesion, cell migration, and matrix remodeling. Cytokines TGF-β, PDGF, and TNF-α affect integrin patterns. Fibroblasts change their integrin patterns during wound healing. Those that facilitate cell migration predominate initially, with those associated with matrix formation and cell attachment increasing in the later stages of healing. The complexity of the system is illustrated by integrin $\alpha^5 \beta^1$, which, in fibroblasts, induces collagenase expression when it is blocked, but in keratinocytes, stimulates collagenase expression when it is active.

The function of collagen in the wound is threefold: wound structure, wound strength, and formation of a matrix that facilitates cellular motility within the wound. This collagen matrix replaces the fibrin scaffold present in the wound during the hemostatic phase of healing. There are 19 types of collagen described in human tissues. Type I, the most common and present in all tissues, composes up to 90% of the collagen in skin. Type III collagen is dominant in the embryo and provides between 10 and 20% of the collagen of adult skin. Type V is predominant in smooth muscle and present during the early phase of wound healing. Types II and XI are dominant in cartilage. Type IV is a characteristic of basement membranes. Type VII is the anchoring fibril of the epidermal basement membrane. Diseases associated with excess or deficiencies of these collagen types include keloid scar formation, Peyronie's disease, Dupuytren's palmar contracture, epidermolysis bullosa, Ehlers-Danlos syndrome, pseudoxanthoma elasticum, cutis laxa, progeria, Werner's syndrome, Marfan's syndrome, and osteogenesis imperfecta types I and II.

The extracellular matrix of the healing wound contains, in addition to collagen, a mixed bag of other components that play an important role: proteoglycans, fibronectin, and elastin. The proteoglycans are secreted by the fibroblast and are composed of a protein core covalently linked to one of several GAGs. These proteoglycans include chondroitin 4-sulfate; dermatan sulfate; heparin and heparin sulfate, keratan sulfate, and hyaluronic acid. Their roles are not completely understood but seem to be important at different stages of proliferation. Hyaluronate is synthesized in large quantities in the earliest postwounding stage and seems to play an important role in cellular migration. By day 5, chondroitin 4-sulfate and dermatan sulfate increase as hyaluronate is degraded. This correlates with increases in tensile strength, suggesting that they play a role in collagen cross-linking and the formation of collagen fibrils. Another component of the extracellular matrix, fibronectin, is the primary component of the fibrin matrix, where it facilitates the migration of the inflammatory cells and later the fibroblasts. It also is important in the facilitation of epithelial migration and endothelial cell migration during neoangiogenesis.

The complex interrelationships of the matrix, cells, cytokines, and growth factors involved in the proliferative phase of wound healing can be summarized by understanding that the macrophage and platelets provide growth factors and cytokines that serve to activate fibroblasts. The fibroblasts in turn secrete the proteins, collagen, and proteoglycans that make up the secondary matrix structure, replacing the initial hemostatic fibrin scaffolding. The matrix itself, through all the growth factors and cytokines in the extracellular milieu, regulates fibroblast migration and proliferation, thus orchestrating the outcome of wound healing.

Fig. 2–2 demonstrates the relationship between time, matrix substances, and the stages of wound healing, as well as the rate of increase of wound-breaking strength. Tensile strength and wound-breaking strength are two different means of accessing the integrity, or holding power, of incisional wounds. Tensile strength is a measurement of load capacity per unit area, and breaking strength refers to the force required to separate the wound edges, regardless of the dimensions. Therefore, depending on the skin thickness, breaking strength can vary widely, whereas tensile strength is constant for wounds of similar length and size. Measurements of these values are used to track influences on wound healing and provide guidelines that can be used to determine such things as suture removal or increase in activity at points in time where adverse

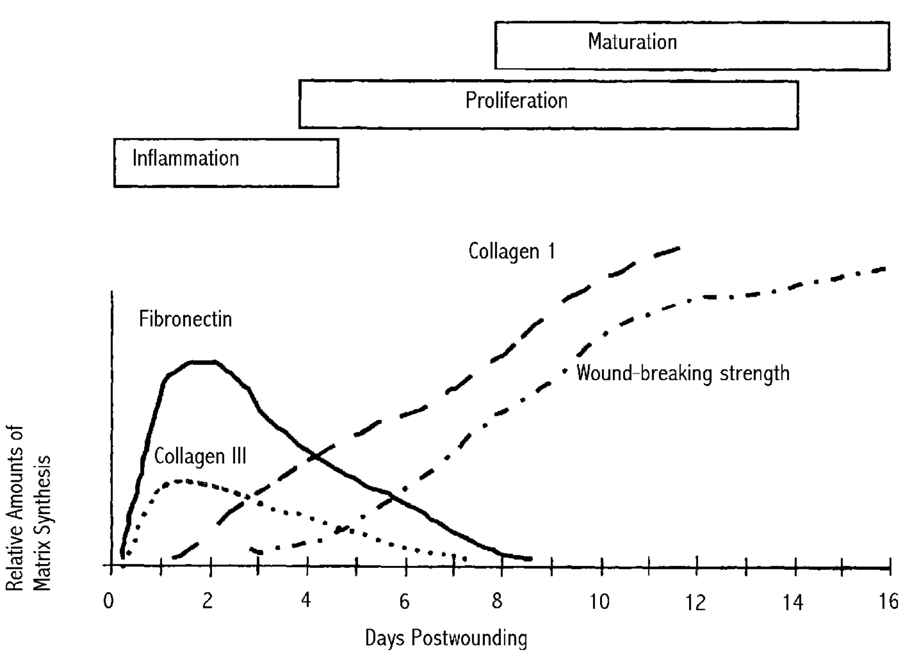


Figure 2–2 *The relationships between the synthesis of extracellular matrix molecules at the wound site, wound healing stages, and wound-breaking strength. Note that the production of type I collagen correlates with wound strength.*

effects on wound integrity can be avoided. All wounds gain strength rapidly from days 14 to 21, creating a sigmoid curve that levels off thereafter. Peak tensile strength is achieved by 60 days after injury, never exceeding 80% of normal skin even with optimal conditions.

In situations where primary closure is not possible or not advised, wounds are left open to heal by secondary intention. In these situations of prolonged reepithelialization, granulation tissue forms, and the numbers of myofibroblasts increase, resulting in significant wound contracture. Wound closure by contraction proceeds by several mechanisms. The theories regarding these mechanisms offer competing views. In one view, the myofibroblast (first described by Gabbiani in 1971), through its "muscular" action within the matrix, pulls the wound edges together. In another, the migration of epithelial cells and the surrounding edges of the wound push toward the middle. And in yet another, it is postulated that fibroblast activity (as discussed by Ehrlich in 1988) in the matrix itself reorganizes the matrix, which pulls in the edges independently of the myofibroblast. The rate of wound contracture depends on the laxity of the surrounding skin, with contraction proceeding faster on the cheek, for example, than on the scalp. The average rate of wound contracture is 0.6 to 0.75 mm per day. Regardless of the forces responsible for contracture, it is a cellular process not related to collagen synthesis. Radiation and cytotoxic drugs inhibit wound contraction, while TGF-β and topical fibroblast growth factor (bFGF) can stimulate contraction. Splints, topical dressings, scabs, early flap or skin graft reconstruction can impede but not prevent closure through contraction. Wound healing via contracture may be ideal in certain circumstances such as contaminated wounds with extensive defects, but the resulting scar, the associated contracture deformity, and the length of time required to achieve closure in large wounds make it less preferable when the means of primary or delayed secondary closure are available in most situations

Primary closure of the skin, or reepithelialization, reestablishes the barrier function of the skin. In primarily closed wounds, contracture does occur, but to a much lesser extent than in the open wound. Meticulous wound closure with eversion of skin edges can counteract the forces of contraction. In an incisional wound, this epithelialization is completed within 24 to 48 hours. Delayed reepithelialization as occurs in open wounds where wound edges are not reapproximated, partial thickness burns, abrasions, and so on, proceeds from the wound edges, from residual rests of epithelial cells in the glandular elements of the dermis (sweat glands and sebaceous glands), and subdermis (hair follicle). Epithelialization proceeds within hours of injury, with the epidermal basal cell's detachment from the basement membrane, migration along the collagen fibers, proliferation of migrating cells forming a monolayer as they leapfrog across the wound, and differentiation once contact with other migrating cells has reestablished contact inhibition. The elongated monolayered cells then differentiate into the more cuboidal basal cells, forming a multilayered epidermis, with a basement membrane, and rekeratinization of the surface cells. The migration and proliferation of these cells are regulated by a variety of cytokines produced by the epithelial cells themselves, by inflammatory cells, and by the underlying mesenchymal cells of the wound matrix. Some cytokines, such as TGF-α, stimulate both proliferation and migration; others, such as heparin-binding epidermal growth factor (HB-EGF) and FGF, keratinocyte growth factor (KFG), and bFGF, stimulate proliferation, whereas TGF-β stimulates migration only.

The result of prolonged reepithelialization is a closed wound with characteristics quite different from normal skin. The regenerated epidermis is thinner, with fewer basal cells, a less well-defined relationship between the dermis and epidermis, and an absence of rete pegs. If the wound involved the full thickness of skin to begin with, there will be an absence of dermal elements as well. The lack of sebaceous and sweat glands, hair follicles, melanocytes, and altered neural endings results in a scar that is dry, painful or itchy, hypopigmented, unattractive, and susceptible to infection, sunburn, and trauma.

The Remodeling Phase

The final stage of wound healing, remodeling begins ~3 weeks after injury and continues for up to 6 months in normal healing, and 2 years or more if abnormal scar formation ensues. A balance between collagen production and degradation begins to form. Apoptosis of endothelial cells is evident, and myofibroblasts disappear. Fibroblasts, in addition to collagen production, synthesize components of the ECM, matrix metalloproteinases (MMPs), elastin, and proteoglycans. Cytokines involved in this process include TNF-α which acts to inhibit TGF-β; fibronectin is broken down, and the cellularity of the injury site decreases. The roles of several of the MMPs have been identified, because they cleave collagen and proteoglycans. Calcium and zinc are important cofactors in the activity of MMPs. Collagen formation begins to balance between production and

TABLE 2–5 PATHOLOGICAL SCAR FORMATION NOTED BY TISSUE TYPES*

Tissue	Excessive Scarring	Insufficient Scar Formation
Skin	Hypertrophic scars	Chronic open wounds
	Keloid	Wound dehiscence
	Scar contracture	Widened, depressed scar
Intestines	Stricture	Fistula
		Anastomotic leaks
Bone	Malunion	Nonunion
	Heterotopic bone	Malunion
Tendons	Adhesions	Rupture
Blood vessels	Atherosclerosis	Pseudoaneurysms
	Stricture	Aneurysm
	Hyperplasia	
Solid organs	Cirrhosis	
	Nephrosclerosis	
Joints	Ankylosis	Subluxation
	Arthritis	

*Different tissues respond with scar formation that can create pathological conditions for the organ. Some of these conditions, associated with either excessive or deficient scar formation, are listed here.

degradation, and reepithelialization of the skin wound is completed. Despite the fact that collagen content is maximal at this stage, the bursting strength of the wound is only 15% of that of normal skin. Collagen cross-linking and the formation of thicker bundles forming fibers correlate with increasing wound tensile strength. The number of intra- and intermolecular cross-links between collagen fibers increases significantly, contributing to a dramatic rise in wound-breaking strength. Although this complex process continues for months or years, it never achieves a stability that matches that of uninjured tissues. Collagen bundles remain disorganized, and the tensile strength never exceeds 80% of that of uninjured tissues.

SCARS

All healing in mammals, except fetal wound healing, proceeds through scar formation. The quality of the scar and its acceptance by the patient and the surgeon depend on a multitude of factors, as we have seen. Clinical discussions of scars should be detailed with the patient, explaining the time course and nature of the process. The wound healing literature has distinguished between normal and abnormal scar formation by characterizing scars as either satisfactory or unsatisfactory. Even a satisfactory result leaves some scar, which may or may not be visible. The first part of our discussion will focus on skin scar formation and its treatment. Discussion of wound healing in other tissues, including bone, cartilage, and the tympanic membrane, will follow. Scars are a consequence of healing in all species and tissues that do not regenerate following injury. Favorable (satisfactory) scars do not cause significant disruption of form or function of the involved tissue. However, even wounds appropriately repaired may have impaired wound healing due to the various conditions listed in **Table 2–5**.

SATISFACTORY SCARS

The favorable scar requires 1 to 2 years to reach maturity. Patient expectations sometimes exceed realistic goals of the scar result, especially early in the postoperative period. Thus the patient should be counseled of this situation preoperatively, if possible. Treatment of the scar during the maturation process with modalities including topical vitamin E, silicone sheet application, scar massage, and intralesional steroid injections may result in more rapid improvement in scar appearance, but in many cases it will have limited effects on long-term results.

Preplanning incisions and repairing injury with attention to the lines of least skin tension will result in satisfactory healing most of the time. Relaxed skin tension lines (RSTLs), also known as the lines of least skin tension, are those lines within the skin of the face (and elsewhere in the body) found at right angles to the direction of underlying facial muscle tension. During the aging process, these lines appear as facial wrinkles. The most critical aspect of facial wound healing is the surgical technique used in relation to these RSTLs. Any wound crossing an RSTL causes potential unsatisfactory wound healing difficulty by several mechanisms. Scars crossing the RSTLs are more likely to be under more tension, creating a tendency to hypertrophy, and more visible, because they lose the natural camouflage of the RSTLs. Contraction of such scars may also result in deformity of adjacent structures, such as the eyelids or the mouth. Proper planning of incisions to avoid crossing the RSTLs will minimize wound-healing difficulties and result in less visible scar formation. Scar revision techniques such as Z-plasty can reorient scars closer to the RSTLs. There are many techniques for maximizing the quality of wound healing while decreasing the chance of wound dehiscence and scar formation: (1) proper planning of the incision; (2) closure of multiple tissue layers to keep tension subepidermal; (3) debridement of irregular and contaminated edges; (4) advancement flap, and rotation flap reconstruction of soft tissue defects, if possible using adjacent matched tissue types, rather than skin grafts; (5) use of nonreactive suture material; and (6) long-term follow-up of the healing process to allow timely intervention if signs of hypertrophic scar, keloid, or contracture appear.

Unsatisfactory Scars

Unsatisfactory scars may be hypertrophic, keloid, or widened. These are scars in which the healing process has been adversely affected. A partial list of such forces includes: (1) nature of the injury (crush, foreign body, severe burn, poor wound management, infection, significant tissue loss); (2) location of the injury (running perpendicular to the lines of least skin tension or Langer's lines); and (3) natural biology of the patient's healing process (previous history of keloid, family history of keloids, or history of other connective tissue abnormalities, malnutrition, previous radiation, chronic hypoxia, current smoking, etc.).

Widened Scars

Widened scars occur during the final phase of wound healing, remodeling, when tension or excess mobility of the wound results in a flat and frequently depressed scar. These scars frequently appear on the shoulder, knee, and back. They are also noted as stretch marks, widening subepidermal disruptions of the dermis that follow changes in subcutaneous skin volume (i.e., pregnancy, weight loss, and accompanying chronic steroid use) or Cushing's disease.

Keloids and Hypertrophic Scars

Keloids and hypertrophic scars are unique human dermal fibroproliferative disorders that develop following trauma, surgery, and burns, or in wounds with significant inflammation. Hypertrophic scarring and keloid formation represent aberrations in cell migration and proliferation, inflammation, increased synthesis, secretion of cytokines and ECM proteins, and remodeling of the newly synthesized matrix resulting in a disordered collagen pattern. Such scars will benefit from either early revision or manipulation of the ongoing healing process by any of several measures: reexcision and rerepair (as might be necessary if the vermilion border is misaligned), local steroid injection, and pressure therapy, among others.

Hypertrophic scars are often mistaken for keloids. They are characterized by thickening, widening, erythema, pruritis, and fibrosis. Such scars are confined to the area of the original wound and are often associated with contractures. Contracture (shortening of the scar) may be associated with deformity, causing loss of function, cosmetic distortion, or impairment in range of motion. In contrast, contraction is a functional part of wound healing that occurs in an open wound, reducing the wound surface as the contraction proceeds. All contracture is caused by contraction, but not all contraction results in contracture.

Hypertrophic scars may be differentiated from keloids by their natural history, in that they tend to improve spontaneously in ~2 years. Aggressive management of hypertrophic scars will force their improvement in a shorter time period. See **Figs. 2–3** to **2–5** for illustration. **Fig. 2–3** demonstrates an auricular keloid that commonly occurs at piercing sites.

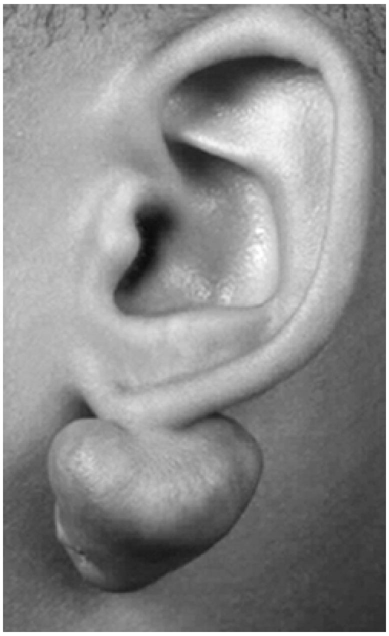


Figure 2–3 *A keloid scar of the pinna after ear piercing.*

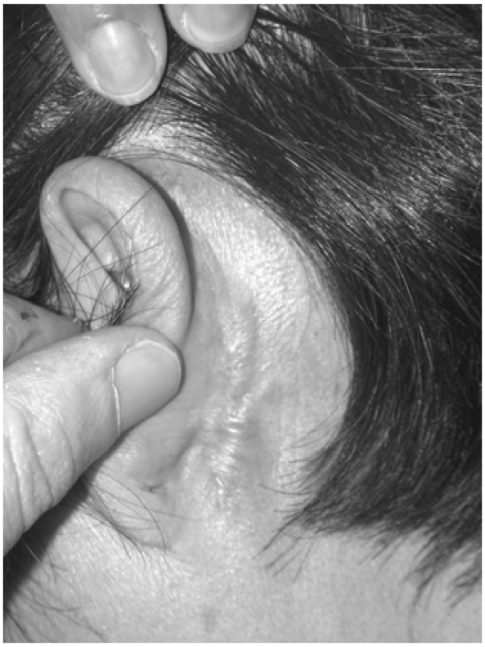


Figure 2–4 *A hypertrophic scar in a postauricular incision for a tympanomastoidectomy.*

Fig. 2–4 shows a hypertrophic postauricular from a tympanomastoidectomy. The scar is over 1 year old and has not been treated yet with any modality. **Fig. 2–5** is a surgical neck scar. Notice how one portion of the scar is hypertrophic, while the scar extending from the lower lip to the chin is much finer, thinner, and less noticeable. It is not uncommon for young adults and teenagers to develop hypertrophic scars in areas that would otherwise heal well for adults or young children.

Alibert in 1806 coined the term *keloid* in a paper differentiating keloids from malignant growths. A keloid is an overgrowth of scar tissue that looks much like a hypertrophic scar, thick, red, and unsightly, usually developing after healing of a skin injury but occasionally occurring spontaneously. This scar tissue extends beyond the borders of the original wound and does not regress spontaneously. Keloids have a strong familial disposition, affect both sexes equally, and may be age dependent, occurring more frequently in adolescence and young adulthood. The genetics of keloids are reported as either autosomal dominant or recessive and are associated with human leukocyte antigen (HLA) factors B14, B21, BW16, BW 35 DR5, DQW3, and blood group A.

Figure 2–5 *Hypertrophic and normal scars on the neck of an adolescent.*

Keloids are often reported to occur more frequently in dark-pigmented skin. Images with keloids appear in western Nigeria Yoruba sculptures of the 13th century. These scars are principally limited to the dermis but can occur in the cornea as well. The occurrence of keloids may be quite variable. Individuals with multiple injuries may develop the scar in some but not all of the wounds. High-risk areas include the chest, in a triangle extending from the clavicle to the lower sternum, ears, arms, and upper back.

The biochemical composition of keloids has been extensively studied and shows significant aberrations. In comparison with normal skin, keloids have increases in water, calcium, histamine, acid phosphatase, alanine transaminase, lactic dehydrogenase, the α globulins, fibronectins, elastin, glycosaminoglycans, chondroitin sulfate, type VI collagen, TGF-β, and abnormal collagen cross-linkage. They contain decreases in procollagen polypeptides due to increased degradation. Collagen types I and III may be increased, decreased, or similar to normal skin. Serum immunglobulins (Ig)M and IgG may be increased or the same as normal. Autoimmune antifibroblast antibodies (AFAs) are found in lymphocyte isolates of patients with keloids. Antinuclear antibodies (ANAs) directed against fibroblasts are found in patients with keloids but not in those with hypertrophic scars. Cytokine production is also abnormal, with increased production of TNF-α, interferon-β, and interleukin-6, and decreased production of interferon-α, interferon-φ, and TNF-β.

Treatment of keloids begins with identification of risk factors, such as family history, examination of old scars, and an understanding that keloids may be region specific (i.e., more likely to occur on the ear, the mandibular margin, and the triangle between the shoulders and the xyphoid on the anterior chest wall). Careful incisional placement, following Langer's lines, meticulous tissue handling, closure with attention to alignment of skin margins, and use of low reactivity suture material, will minimize the inflammatory process and initiate the best possible circumstances for healing. But a keloid may appear despite preventive management.

Keloid formation responds to intervention only a certain percentage of the time, so the initial conversation with the patient must include a detailed education. Large disfiguring keloids may require initial surgical management in combination with other modalities. Smaller lesions may respond to triamcinolone acetonide injections, pressure (such as the use of broad-backed clips on earrings for lobe keloids), or 20% hydrocortisone topical applications. Recurrence rates are reported from 5 to 50% following an initial successful response. With excision and immediate local steroid treatment, topical silicone, and pressure dressing, ~50% successful response rates can be anticipated. Excision and topical treatment accompanied by immediate radiation therapy (short course, 3–50 Gy) results in a 20 to 45% recurrence rate. Brachytherapy is now available, and initial reports of 20% recurrence in 1 year have been reported. Brachytherapy in conjunction with hyaluronidase injection in one study has been reported to have a zero recurrence rate at 1 year. Various regimens employing laser therapy, including the carbon dioxide laser at 10,600 nm wavelength, neodymium:yttrium-aluminum-garnet (Nd:YAG) laser at 1064 nm, and the argon laser at 488 nm, all report recurrence rates ranging from 25 to 90%. Repeated treatment with cryotherapy yields up to a 75% improvement response rate, and in combination with intralesional corticosteroid injection, prevents recurrence more significantly than with either alone.

Advances in understanding the biochemistry of wound healing have created new options for keloid prevention and treatment. Interferons-α and -β, and -γ reduce fibroblast synthesis of collagen types III, VI, and I and increase collagenase activity. Interferons-α and -β also reduce glycosaminoglycan synthesis. These actions have not translated into clinical efficacy as yet, and repeated injections into active keloids reduced keloids by 50% only. Injection into surgical excision sites has had limited success. The effective actions of anti-TGF-β, procollagen peptides, *cis*-hydroxyproline, and pentoxifylline are still under investigation.

The list of previously investigated failed therapies is even longer than the list of those currently accepted as at least partially effective: injections of formalin, mustine combined with triamcinolone, pepsin, hydrochloric acid, creosote, nitrogen mustard, thiotrexate, topical ultrasound, surgery combined with tocophol, thiotepa, penicillamine, oral betaaminopropionitrile (BAPN), colchicine, topical zinc tape, retinoic acid, and tetrahydroquinone. Many of these agents, although known to play a significant role in collagen synthesis and degradation, have failed to prevent keloids, despite excellent rationalization of their possible efficacy.

FAILURES OF WOUND HEALING

When a wound does not heal, consideration of the intrinsic or extrinsic factors contributing to the problem should nearly always yield an appropriate solution. Extrinsic factors such as the cachexia of advanced malignancy and malnutrition; systemic diseases such as diabetes, hypertension, arteriosclerosis, and sickle cell disease; autoimmune conditions; neurological conditions such as multiple sclerosis and spinal cord injury; pulmonary insufficiency; alcoholism; smoking; and steroid use must be addressed in advance of surgery for a successful outcome to be ensured. Optimizing those circumstances that are not completely reversible may require several weeks of advanced preparation. In the presurgical and postoperative phases, attention to these factors by close monitoring of the patient through clinical observation and laboratory analysis is mandatory. Protein malnutrition, morbid obesity, trace mineral deficiency, hypoxia, hyperglycemia, smoking or chewing tobacco, steroids, irradiation, hypothyroidism, and use of cytotoxic drugs such as methotrexate negatively affect wound healing by their actions, inhibiting the inflammation and proliferation required to initiate the appropriate cellular and biochemical responses of wound healing. Failure to note and address these adverse conditions will harm both the patient and the reputation of the physician/surgeon.

Intrinsic conditions that can affect wound healing include previous radiation in the wound site, infection, hematoma and seroma, ischemia, vasculitis, retained foreign body, excessive motion across or tension on the suture line, and persistence of tumor following resection. Social circumstances of the patient and living conditions such as sanitation, domestic violence, and alcohol or drug abuse should also be considered. Intractable persistent wounds should only be considered factitious when all other possibilities are ruled out. Wound infection remains one of the most common and expensive complications of modern surgery. Prediction of risk of surgical wound infection includes such factors as surgical time in excess of 2 hours, hypothermia, hypoxia, contamination, and the presence of multisystem diseases. Risk reduction strategies for the perioperative period include (1) maintenance of oxygen above 90%

for at least 2 hours following surgery, (2) maintenance of body temperature in the normal range during and after surgery, (3) use of preoperative antibiotics if contamination of the wound is predicted, (4) a normal wound oxygen tension, (5) an adequate intravascular volume, (6) albumin level above 3%, (7) adequate pain control, and (8) appropriate dressing techniques.

SPECIFIC NUTRITIONAL FACTORS AND THEIR EFFECT ON WOUND HEALING

Optimum healing occurs only in the presence of optimum circumstances. It is well known that grossly malnourished patients experience delayed wound healing and increased rates of wound infection. These factors can be reversed. The marked increase in energy demands and the caloric requirements that can lead to cannibalization of lean body mass contribute significantly to morbidity. Several clinical studies have demonstrated that malnourished patients whose caloric needs are met during the week before surgery can exhibit normal wound healing capacity. In addition to overt malnutrition, metabolic deficiencies or disorders of specific nutrients can result in delays in wound closure. These range from macronutrients such as amino acids to micronutrients including vitamins and trace minerals. Attention to the nutritional status of the patients is an important component of presurgical planning. The head and neck cancer patient has a strong disposition toward malnutrition, and subsequently poor wound healing. There are several reasons why this is so. Because of odnophagia or dysphagia, these patients have poor oral intake of foods. The high prevalence of alcoholism and an unbalanced diet in the head and neck cancer patient population contribute to poor baseline nutrition. Also, hypermetabolism of the tumor may cannibalize the host patient.

Optimum nutritional support for the compromised patient has several key components. First, an assessment of the caloric requirements needed to meet energy demands and a provision of sufficient protein for routine protein synthesis including calculated requirements for wound healing should be part of the patient's evaluation. Second, dietary supplements with sufficient micronutrients should be initiated and maintained through either the oral or the intravenous route. Third, consideration to the use of anabolic hormones should be given in those cases where preexisting malnutrition and a prolonged recovery can be anticipated.

Amino Acids

Gross deficiency of dietary amino acids and protein results in delayed wound closure through impairment of protein (mainly collagen) synthesis. Protein breakdown occurs during stress-induced catabolism, and unless adequate protein replacement is provided during this process, a rapid loss of lean body tissue including muscle and visceral proteins will continue even if carbohydrate-based calories are being provided. A cessation of caloric protein intake can actually result in the complete halting of collagen synthesis. Decreased protein intake also impairs the immune response and inflammation. Supplements beyond normal requirements are not beneficial, except arginine, which has strong evidence that enteral or parenteral supplementation beyond regular dietary requirements results in accelerated wound healing and enhancement of early collagen deposition. In protein malnutrition, methionine (one of the essential amino acids) is converted to cysteine and is crucial to the inflammatory process and the production of fibroblasts. Loss of lean body mass contributes substantially to such postoperative complications as profound weakness, a significant decrease of immune function contributing to wound infection and ammonia and other sources of sepsis, the already described impairments of wound healing, and increased risk of pressure sores. An inadequate intake of protein and calories to meet demand results in a condition known as protein energy malnutrition (PEM).

Vitamin A

Vitamin A is an essential fat-soluble vitamin. Its primary physiological roles include maintaining epithelial integrity and cell membrane integrity, serving as a protective factor against infection, and operating as a cofactor in collagen synthesis. It serves an important role in the wound healing process at multiple levels. Deficiencies in vitamin A result in decreased collagen synthesis and decreased rates of epithelialization. Serious injury and systemic stress (e.g., sepsis) increase vitamin A requirements and thus predispose to deficiency and impaired wound healing. Multiple studies have demonstrated that supplemental vitamin A can reverse the inhibitory effects of systemic corticosteroids, can potentially reverse the negative wound-healing effects of diabetes, cyclophosphamide, and radiation, and has multiple beneficial effects on wound healing, including an increased inflammatory response, increased collagen synthesis, increased macrophage infiltration of the wound site, and increased lability of the lysosomal membranes of inflammatory cells. Supplemental vitamin A also appears to decrease the incidence and severity of stress ulcers in both experimental animal models and in human patients.

Vitamin A is stored in the liver. Uncomplicated elective surgical procedures will not result in vitamin deficiencies. Prolonged decreased food intake, malabsorption, and chronic malnutrition will result in vitamin A deficiencies. All patients experiencing severe stress, such as a major burn injury, multisystem trauma, or postoperative complications, should receive 25,000 international units (IU) per day. This is 5 times the minimum daily requirements but should not result in significant toxicity. Vitamin A should also be given to those patients who are taking chronic steroids. Vitamin A given orally at 25,000 IU per day or topically as an ointment as 200,000 IU every 8 hours, will reverse the inhibiting effect steroids have on wound healing. Larger doses should be avoided, because vitamin A is toxic to the liver and cornea.

Vitamin C (Ascorbate)

Sixteenth-century explorers and physicians clearly described the condition now known as scurvy. Long-term deficiency of vitamin C (at least 6 months of ascorbate-free diet) has long been known to delay wound healing and can actually cause previously healed wounds to reopen. In the 18th century Lind demonstrated that lime-juice could prevent scurvy. This made it possible for the British to deliver healthy soldiers and sailors after long sea voyages and gave Britain the manpower to create the empire upon which the sun never set; hence the nickname “Limey” was given to British sailors. In 1926 Wolbach published a classic study. He created an experimental scurvy in guinea pigs, which demonstrated the failure of collagen synthesis. This effect was immediately reversed once ascorbic acid was given. Vitamin C is a well-established cofactor in the hydroxylation of the amino acid proline to hydroxyproline, an amino acid required in collagen synthesis. Vitamin C deficiency manifests itself as a failure of cross-linking of collagen fibers, and subsequently a loss of wound strength. Other cofactors in the hydroxylation process include oxygen, α-ketoglutarate, and ferrous iron.

Cellular effects of vitamin C deficiency in the healing wound include the proliferation of fibroblasts that are immature and do not mature, and the formation of defective capillaries that can cause local hemorrhage. Biochemical effects include the failure of formation of mature extracellular collagen, the production of chemically detectable levels of alkaline phosphatase, and proteolysis of intracellular unhydroxylated protein.

Vitamin C deficiency is also associated with an increased rate of wound infection. Impaired collagen synthesis reduces the ability of the body to wall off an abscess because of reduced collagen synthesis. There is also a reduction in neutrophil function, resulting in decreased bacterial killing. Ascorbic acid is also involved in the reduction of oxygen to superoxide. Complement-dependent immune reactions are seen to be depressed. Popular belief that vitamin C protects against the common cold or shortens its normal course has not been confirmed.

Excess doses of vitamin C do not enhance normal wound healing, and megadoses may cause renal oxalate stones. Replacement during extended illness or periods of extreme physiological stress maintains normal healing, so ascorbic acid should be administered in 1 or 2 g doses on a daily basis until recovery is complete. This daily dose is the equivalent of the total amount of vitamin C stored in the adult body, a lack of reserve that explains the importance of replacement and maintenance.

Vitamin B Complex

The B-complex vitamins serve as cofactors in a wide variety of enzyme systems. Deficiency in these vitamins, especially pyridoxine, pantothenic acid, and folic acid, results in major effects on resistance to infection. Antibody formation and neutrophil function are impaired in deficiency states. Supplemental intake of 5 to 10 times the minimum daily requirements of these vitamins is recommended in the treatment of severe injury and acute illness. Estrogens are known to increase requirements for pyridoxine and for folic acid. Addition of folic acid to prenatal vitamins significantly reduces neural tube defects like meningomyelocele.

Vitamin D

Vitamin D is essential to the normal absorption, transportation, and metabolism of calcium, and indirectly for phosphorus metabolism. It is important in normal bone growth and in bone healing. Severe vitamin D deficiency results in rickets that can be avoided through moderate sun exposure. Recent work on hospitalized elderly patients in the northeastern United States recognized that over 50% of hospitalized elderly are vitamin D deficient. With increased avoidance of sun exposure and reduced dairy intake, this problem is increasingly widespread. Vitamin D deficiency results in loss of bone strength, increases the risk of low-impact fractures, and retards the healing of bone, when fractured, accidentally or electively.

Vitamin K

Vitamin K is an essential cofactor in the synthesis of prothrombin, and the clotting factors VII, IX, and X are required for the synthesis of calcium-binding protein. In vitamin K deficiency a bleeding diathesis occurs. Parenteral injections of vitamin K can be used to reverse the effects of Coumadin (warfarin sodium). In the presence of liver disease vitamin K may not be able to promote synthesis of adequate amounts of prothrombin.

Vitamin E

Unlike other vitamins, large doses of vitamin E have been found to inhibit healing, decreasing tensile strength and reducing collagen accumulation. Vitamin E plays an important role in membrane stabilization, neutralizes lipid peroxidation, and limits the levels of free radicals, peroxidases, and other products of lipid peroxidation. This antioxidant activity has been promoted for its antitumor and antiaging effects, but recent studies have not demonstrated a statistical benefit. It is known that supplemental vitamin E increases the risk of bleeding in the perioperative period. Oral supplements should be stopped 1 to 3 weeks before planned elective surgery.

Minerals

Iron

Like vitamin C, iron is required for proline and lysine hydroxylation. It is also a critical cofactor in the replication of deoxyribonucleic acid (DNA). Iron deficiency anemia will diminish oxygen transport, an effect that will impair oxygen delivery, causing a secondary effect on wound healing and on bacterial killing.

Calcium and Magnesium

Calcium and magnesium are cofactors in protein synthesis and in the function of the collagenase enzyme, which is useful in the remodeling phase of wound healing. Calcium supplements should be considered for all women in the perimenopausal period and for individuals who avoid dairy products. Carbonated beverages impede calcium absorption through their high phosphorus content and may impair bone formation during adolescence and young adulthood, contributing to osteoporosis in later years. Calcium supplements, with vitamin D when appropriate, will facilitate bone healing in the deficient state but are not known to promote healing beyond normal when nutritional status is adequate.

Zinc

Zinc functions as a cofactor for over 70 known enzymes, including ribonucleic acid (RNA) and DNA polymerases and other transferases. Therefore, mitosis is impaired, and the cell proliferation required for normal healing is decreased. Deficiency of zinc results in decreased fibroblast proliferation, decreased collagen synthesis, and delayed epithelialization of wounds. Zinc deficiency is not common, but it may occur in patients with large burns, diffuse sweating, severe trauma, chronic alcoholism, intestinal fistulas, and cirrhosis. Chronic zinc deficiency has been reported in Middle Eastern children and is characterized by short stature, mild anemia, and hypogonadism. Supplemental zinc will not increase wound healing unless zinc deficiency exists.

ROLE OF OXYGEN

No discussion of wound healing would be complete without a review of the role oxygen plays in the process. Oxygen delivery to the healing wound is controlled by multiple factors. Inspired oxygen is transported across the alveolar capillary gradient of the lung, is bound to hemoglobin delivered by the circulatory system, then transmitted across the capillary bed, where it dissolves in the extracellular tissues and thus reaches the point of injury. Respiratory insufficiency, anemia, cardiac insufficiency, peripheral vascular insufficiency, and defects in hemoglobin (including anemia, carboxyhemoglobin, and sickle cell disease) will all adversely affect the healing process. A failure of oxygen delivery is one of the common pathways of the wound-healing impairment associated with diabetes mellitus, irradiation, arteriosclerosis, and chronic infection. Conversely, the rate of healing is a function of arterial oxygen tension through a certain physiological range. Assuming normal oxygen delivery mechanisms are in place, the oxygen tension distributed throughout the wound varies directly with the proximity of the capillary. Close to the wound capillary oxygen tension ranges between 60 and 90 mm Hg. At 150 υ away from the capillary, oxygen tension approaches zero. Actively dividing fibroblasts are found in close proximity to the capillary, whereas macrophages are the only cells found at the distance where oxygen tension approaches zero. Collagen synthesis requires a partial pressure of oxygen (PaO_2) $>$ 40 mm Hg pressure. Oxygen is used not only in the energy-requiring processes of protein synthesis and cell replication but also in the biochemical hydroxylation of proline and lysine molecules.

Hyperbaric oxygen (HBO) has been used to increase oxygen tensions in hypoxic, difficult wounds. Treatment of wounds with HBO stimulates fibroblast proliferation and differentiation, increased collagen deposition, and cross-linking, neovascularization, and microbial lysis. Limited clinical data suggest that this modality can be successful in enhancing the healing of poorly healing irradiated or diabetic wounds.

OTHER FACTORS INFLUENCING WOUND HEALING

Smoking

Smoking has several adverse effects on primary wound healing, as well as multiple effects on systemic well-being. Long-term consequences including emphysema, peripheral vascular insufficiency, and carcinogenesis significantly increase the risks of elective or reconstructive surgery. Inhalation of the combustion products of tobacco includes carcinogenic tars, carbon monoxide, and nicotine as the active toxic ingredients.

Nicotine is a stimulant and the primary addicting agent in smoking. Metabolic effects of nicotine include increased platelet adhesion and increased thrombus formation within the microvasculature. Nicotine also directly inhibits keratinocyte migration, prolonging reepithelialization, and inhibits proliferation of red blood cells, macrophages, and fibroblasts, further impairing wound healing.

Associated effects of smoking cause peripheral vasoconstriction, reduce capillary perfusion, increase bleeding, and through carbon monoxide binding shift the oxyhemoglobin curve to the left, reducing oxygen delivery. Hydrogen cyanide is another product of combustion in smoking. The deleterious effect of hydrogen cyanide is on oxidative metabolism and oxygen transport at the cellular level, interfering with cellular respiration. Smoking directly increases matrix metalloproteinase-1 messenger ribonucleic acid (mRNA), a product of dermal fibroblasts that degrades collagen. In addition to increasing degradation of collagen, nicotine reduces production of types I and III collagen by as much as 40%. Normally, collagen represents ~70% of the dry weight of the skin. Increased degradation, in combination with reduced production, could be one of the central mechanisms contributing to the thinned, wrinkled skin of smokers' faces.

Smoking is a significant cofactor in osteoporosis, impairing calcium metabolism, reducing fracture healing, and increasing fracture risks. The combination of toxins, causing direct vasoconstriction, intravascular thrombus, collagen disruption, impaired oxyhemoglobin dissociation, and oxygen consumption, affects all aspects of the healing process. The carcinogenic activity of smoking is well known in relation to lung cancer but less recognized as a cofactor in both squamous cell and basal cell cancers of the skin.

Smoking is a direct contraindication to several elective procedures, especially facelift surgery, because the risks of skin necrosis, hematoma formation, and scarring create a high probability of an unsatisfactory outcome. Smoking also increases the complication rate of general anesthesia and the risk of postoperative pneumonia. Studies have indicated that smoking cessation, with complete withdrawal at a minimum of 2 weeks before elective surgery, can reverse many of the acute toxic effects of tobacco. Complete cessation will diminish the risk of malignancy.

Alcohol

The negative effects of alcoholism and its associated malnutrition have already been mentioned. Alcohol withdrawal continues to carry a significant risk of morbidity and mortality. Recognizing the risk in the preoperative period and initiating the appropriate treatment for delirium tremens will reduce the risk of poor outcomes. Supplemental thiamine and treatment with psychotropic drugs (benzodiazepam, Librium, Thorazine, or other recommended agents) should be used as prophylaxis.

Steroids and Anti-Inflammatory Drugs

Steroids inhibit wound macrophages and epithelial cell migration and reduce essential biochemical processes involved in fibrogenesis, angiogenesis, and wound contraction. Vitamin A will reverse the effect of steroids on the inflammatory process and promote epithelial repair. Anti-inflammatory drugs such as aspirin and ibuprofen decrease collagen synthesis by up to 45%. The effect is dose dependent and reverses slowly. Stopping aspirin in the preoperative period should be considered 4 to 6 weeks in advance of elective surgery. The list of drugs and supplements, especially over-the-counter medications containing salicylates, should be reviewed with each patient. The inhibitory effect on the inflammatory process is mediated through prostaglandin. Salicylates are also well known to increase bleeding time and risk of hematoma by their activity in reducing platelet stickiness.

Lathyrogens

Lathyrism was well known among ancient Greeks. Eating the ground pea (*Lathyrus odoratus*) results in loss of collagen in tendons and ligaments, as well as fascia. Poisoning leads to giant hernias, knee instability, and vertebral dislocation. The active ingredient is BAPN. The effect is mediated through blockage of the aldehyde intermediates in collagen cross-linking. The strength of the collagen bundles is lost, and tissue integrity is disrupted. Both BAPN and a related lathyrogen, *d*-penicillamine, have been used as pharmaceutical agents to control scar tissue, to little benefit.

Oxygen-Derived Free Radicals

Radiation, ischemia, inflammation, and certain chemicals release oxygen-derived free radicals that are cytotoxic and highly reactive. They cause cellular injury by degrading hyaluronic acid and collagen, destroying cell membranes, interfering with protein enzyme systems, and disrupting intracellular membranes. Oxygen radical scavengers such as superoxide dismutase (SOD) and allopurinol prevent damage from these radicals by blocking xanthine oxidase (XO). Intracellular generation of free radicals is one of the factors contributing to damage associated with aging, reperfusion syndromes, hyperoxygenation syndromes, chemical-induced tissue injury, drug-induced hemolytic anemia, and vitamins A and D deficiency. Blocking these radicals is the goal of many current antiaging regimens and in the salvage of replanted tissues and failing flaps.

Aging

During World War I, Alexis Carell and coworkers noted the correlation between rates of wound contracture and age. Open wounds among 20-year-old patients closed in 40 days. This rate slowed progressively, rising to 56 days in 30-year-olds and 76 days in 40-year-olds. In the burn literature, significant differences in healing and rates of morbidity and mortality correlate with age. Patients over the age of 50 are considered part of the elderly risk group for complications. Factors believed to contribute to this decrease in the rate of healing include reduction in the amounts of connective tissue deposited over time, reduced hydroxyproline accumulation, decreased cellular activity, including reduced numbers of cells present, reduced secretory products, and significantly reduced DNA and RNA synthesis. Epithelial migration is slower. Age is affected by numerous factors other than simply time. Genetic coding, repair, and control of cell replication and repair are among the cellular factors involved. Reduced tolerance to ischemia is also implicated as one of the major factors. Changes in growth hormone production, estrogen and testosterone levels, and reduced sensitivity to these hormones, among others, result in alterations in body fat/protein ratios, bone density, skin thickness, and vascular reactivity. Although healing does occur, its time scale is altered significantly, and cellular proliferation is reduced in all phases of healing.

Significant features of healing in the aged include (1) reduction in rate of gain of tensile strength and in the absolute value at completion, (2) reduced wound closure rate and bursting strength, (3) delayed start for all cellular events in the course of healing, and (4) reduced cellular proliferation and substance production. Healing in the aged is quantitatively impaired and delayed, but it can be achieved if deficiency states are corrected, general principles of good surgical technique and postoperative management are followed, and associated conditions, such as diabetes, heart disease, and pulmonary insufficiency, are adequately addressed.

Facial skin aging occurs in sun-exposed skin more severely than in non-sun-exposed skin surfaces. Both genetic (intrinsic) and environmental (extrinsic) factors contribute to the changes seen in photodamaged and aged skin. Photoaging is characterized by wrinkles, pigment irregularities, loss of skin tone, loss of resilience, and malignant degeneration. Ultraviolet (UV) radiation increases collagen-degrading MMPs and reduces collagen synthesis. Damage affecting both mitochondrial and genomic DNA affects repair mechanisms in ways that result in loss of collagen and elastin fibers. Furthermore, UV A and B increase elastase activity. Neutrophil production of serine elastase is associated with increased fibronectin breakdown products, which themselves are harmful to tissues. UV damage, causing active cellular and matrix degradation, and chronological aging effects such as reduced cell proliferation and production capacity, are some of the underlying mechanisms of skin aging. The underlying pathogenic agent for much of this damage is believed to be reactive oxygen species that are known to affect growth, differentiation, senescence, and connective tissue degradation. Some of these effects are reversed by topical applications of vitamin A as retinoic acid. Use of effective sunscreen will also protect the skin by inhibiting the production of oxygen radicals. Possible reversal of these age-related effects through manipulation of the wound environment by application of exogenous growth factors has been evaluated extensively but has not yet found a definable role in wound management. Topical cytokines and retinoids have found a role in enhancing skin stability and increasing collagen formation, and may improve healing characteristics, but they are used most frequently at the present time for their antiwrinkle effects.

HEALING IN SPECIALIZED CIRCUMSTANCES

Skin Grafts

Healing of skin grafts proceeds in several stages. Additional processes accompany the usual steps of inflammation, proliferation, and maturation. The skin graft that is used may be a thin ($<$ 0.0015 in. thick) split-thickness graft harvested from a remote donor site, a thicker graft ($>$ 0.0015 in. thick), or a full-thickness skin graft. The thinner the graft harvested from a donor site, the more quickly the donor site will heal, with less pigment change and scar formation at the donor area, and more scarring, contraction, and contracture at the recipient site. Full-thickness skin grafts are frequently taken from areas of high skin laxity, permitting closure of the donor area with primary intention. Full-thickness skin grafts typically have a better color match and contract as healing progresses. It is speculated that the amount of dermis transferred with the epidermis in the graft is an important cofactor in quality healing.

The skin graft, when applied to a recipient bed that is free of contamination and has good vascularity, will be nourished for the first 72 hours by a process known as imbibition. The graft adheres to the bed through proper surgical technique (sutures, bolster tie-over dressing, and bulky dressing immobilization) and naturally occurring fibrin glue. Bacteria in the recipient bed, especially staphylococcal species, produce fibrinolytic enzymes that will prevent adherence of the graft, giving it an appearance of "dissolving" within a few days of application. If the graft is properly immobilized, neovascularization brings capillary blood flow directly into the transplanted dermis and epidermis within 72 hours. Take off of the graft is accomplished when capillary ingrowth, fibrinogen fixation, and cellular proliferation in the graft are established. Transplantation of epidermis alone can be done if cultured epidermal cells are used. Commercial production of autologous epidermal cells is now widely available, and homograft epidermal cells cultured from fetal foreskin are also effective and available in many countries, and under certain circumstances in the United States. Conditions where such techniques may be necessary include massive burns, congenital conditions such as epidermolysis bullosa, and autoimmune chronic skin disorders such as severe toxic epidermolysis necrobioticum. Cultured epidermis is expensive, and current research attempts to find useful applications in chronic skin ulcers, smaller burns, and other such open wounds have yet to be shown to be clinically valuable.

Cartilage

The three-dimensional structure of the ear, nose, larynx, and tracheal airway is supported by cartilage. There are three types of cartilage in humans: (1) hyaline cartilage, found in joints, rib cage, and trachea, functions to dissipate loads on bone and in joints; (2) elastic cartilage occurs in the ear, nose, and larynx, providing the semirigid support required by their structure and function; (3) fibrocartilage, found in the intervertebral disks, tendon attachments, and temporomandibular joint, serves to transfer loads between tendon and bone and to cushion joints.

Cartilage consists of a highly organized matrix of primarily type II collagen and proteoglycans and a cellular component, chondrocytes, embedded in a multilaminated capsule. Cartilage contains no lymphatic or blood vessels. Thus small, localized injury results in no bleeding or inflammatory process, and healing is not initiated unless surrounding soft tissues are also injured. Formation of scar tissue following cartilaginous injury results in disorganized bundles of type I collagen, with the result that the shape and structure of the cartilage can be altered. Initial healing in cartilaginous tissue that has been significantly traumatized is similar to the wound healing process of skin that has already been discussed. A cascade of peptides, growth factors, and inflammatory cells characterizes the initial stage. Where the process of cartilaginous wound healing differs from that of skin is in the acute need for a delicate balance between the deposition of scar tissue and the actual regeneration of cartilage. Inflammation and subsequent type II collagen deposition inhibit the regrowth of cartilaginous tissue. Fibrocartilage, as found on the articular surfaces of the temporomandibular joint, has much better regenerative capacity than hyaline cartilage. Fibrocartilage repair is mediated by mesenchymal cells of the adjacent bone marrow, which have greater regeneration abilities than do mature chondrocytes. Thus, because of the proximity to bone marrow, full-thickness cartilage defects of the joint are replaced by fibrocartilage, and partial-thickness defects are mainly repaired by depositions of fibrous scar tissue. Hyaline cartilage heals with the initial formation of a fibrocartilage intermediary that is eventually replaced by hyaline. The primary mediator of hyaline regeneration is a mesenchymal cell from adjacent marrow, not chondrocytes.

Maintaining the shape of cartilage is a primary consideration in cartilage repair in the ear, nose, and airway. Careful reapproximation of the structure, splinting for extended periods of time, and minimizing inflammation can result in a satisfactory outcome. Infant cartilage can be retrained and reshaped into aesthetically improved positions with only taping and splinting, correcting such deformities as lop ear and cleft lip nasal distortions if initiated in the perinatal period.

Mature cartilage is much more resistant to reshaping by manipulation. The most common complication of cartilage healing is deformation or warping. This is well illustrated in cases of nasal deviation after trauma or corrective nasal surgery. Even after prolonged splinting of the nose or septum, the cartilage has a tendency to redeviate. Cartilage heals with a fibrous union, leaving the structural integrity impaired in ways that are not yet well understood.

Blunt injury to cartilage may result in subperichondral hematoma. Left untreated, this can result in pressure necrosis of the cartilage. In the ear, this results in subsequent fibrosis and "cauliflower" or "boxer's" deformity. In the nose, this can result in septal perforation and loss of dorsal support.

Better understanding of healing cartilage will result in improved care for fractures, inflammatory diseases, and diseases of aging, including osteoarthritis and posttraumatic joint disorders.

Bone

Bone is a complex tissue containing many unique cells: osteoblasts (derived from mesenchymal, osteoprogenitor cells), responsible for bone formation; osteocytes, a type of osteoblast surrounded by bone matrix; and osteoclasts (related to the monocyte/macrophage cell line), which are multinucleated cells that direct bone resorption. Just as the healing process described earlier can be divided into stages, bone heals through a sequence of stages: (1) injury, (2) soft callus, (3) hard callus, and (4) remodeling. Understanding the distinct characteristics of bone healing requires knowledge of the physiology of bone, its matrix, and its unique cells. The connective tissue matrix of bone is both inorganic and organic. The organic matrix is similar to dermis, consisting of collagen type I, glycoproteins, and proteoglycans. The inorganic matrix includes mainly calcium salts, predominantly hydroxyapatite. Bones are composed of 92% solid material. The organic component of the bone matrix consists of 98% collagen and 2% glycosaminoglycans and proteoglycans. Blood supply to the bone is derived from three sources, the periosteum, endosteum, and surrounding soft tissue.

Two specific types of bone exist, membranous and endochondral, both derived from mesenchymal tissues. They are distinguished by the means through which they ossify. Endochondral bone initially forms as cartilaginous tissues at the epiphysis and then ossifies. Membranous bone is formed from preexisting mesenchymal cells that differentiate directly into osteoblasts and that form osteoid material without initial cartilaginous intermediates. Membranous bone includes the skull and most of the facial bones. The long bones of the skeleton are endochondral. Further distinction between bone is made on a morphological basis, cortical or cancellous. Few tissues in the adult human have a better capacity for the regeneration of both form and function than bone. Primary reconstruction by reduction and fixation of fractures will result in the restoration of facial appearance. Fracture or osteotomy stimulates bone regeneration so efficiently that, when completed, the fracture site may be undetectable. Abnormal healing occurs and may result in such exuberant bone formation, malunion if the fragments are misaligned, or nonunion if stable fixation is not achieved. Defects in the facial skeleton do not initiate regenerative growth, and nonunion of improperly immobilized fractures is a common outcome. One reason for the persistence of a defect is an inadequate supply of inducible proteins and growth factors required for osteogenesis, the most important of which is bone morphogeneic protein (BMP). The source of BMP is demineralized bony matrix, something that is deficient in cases of large bony voids. The ingrowth of fibrous tissue from the surrounding connective tissue can mechanically interfere with the process of bony union and impair the healing of bone defects. Finally, a lack of rigid scaffolding across the void inhibits conduction of osteoblasts and osteoclasts into the wound site.

Normal bone healing proceeds through injury and inflammation. Following this initial stage of healing, beginning between 3 and 4 days after injury, bone begins a unique healing process, osteoinduction and osteoconduction, to begin the formation of soft callus. Cells involved in this process include fibroblasts, chondroblasts, and osteoblasts. Mesenchymal cells are transformed into bone cells (osteoinduction) by effects from growth factors, physical signals, and other chemical modulators. Ingrowth of capillaries and osteoprogenitor cells into the fracture site (osteoconduction) arises from adjacent soft tissue and bone, primarily from the periosteum. Migration into the wound by fibroblasts, osteoblasts, and chondroblasts initiates cellular secretions that establish a mixture of fibrous tissue, composed of cartilage and loosely woven bone known as a soft callus. The callus forms a fibrous union surrounding the break and entering the marrow cavity. Type II collagen composes 50% or more of this callus. With formation of the soft callus, the fracture begins to stabilize, and inflammation ends. Movement at the site, bacterial contamination and infection, soft tissue injury around the site, insufficient vascularity, and malalignment can cause failure of soft callus formation.

Hard bone forms in the third stage, as the cartilaginous portion of the soft callus is replaced by endochondral ossification beginning ~3 to 4 weeks after injury. Type II collagen is degraded and replaced by type I. The organic matrix calcifies as calcium hydroxyapatite is deposited and the developing calcified matrix is interconnected by brisk neoangiogenesis. The new blood vessels permit entry of osteoprogenitor cells that then mature into osteoclasts. The hard callus is visible radiographically as an excess of calcified bone in and around the fracture site. The concave surface is electronegative and forms more osteoblasts, while the electropositive charge of the convex surface is populated by more osteoclasts. The presence of electric charges on the convex and concave surfaces of the callus can be influenced at this stage by externally applied electrical stimulation, a technique that may be used to promote healing, especially when delayed union or nonunion is suspected.

The final stage of bone healing, reshaping, involves remodeling and modeling. Remodeling involves the cell-mediated breakdown and reformation of bone, and modeling normalizes the bony macrostructure, providing an orientation reflecting the lines of bone stress. Wolff's law refers to the ability of bone to reorient its microstructure along the lines of biomechanical stress. As these processes occur, continuing over years, radiological evidence of the fracture nearly disappears.

Bone healing in membranous bone differs from that of endochondral bone because healing can proceed without the intermediate stages of soft and hard callus formation. This is called primary bone healing. New bone bridges the fracture site early, with type I collagen predominating. Within 6 weeks of injury new osteons, with well-formed lamellae and haversian canals, are present, providing a regeneration of the bone at the fracture site. Facial fractures need to be reduced and stabilized soon after injury, not only to minimize pain and restore appearance and function, but also because strength in the fracture site may make the reduction impossible within 2 weeks. This type of primary healing in endochondral bone may occur with minimal tissue damage and immediate rigid fixation of the fracture.

Bone Grafts

Bone grafts transplanted into the site generate new bone formation through three processes: (1) the presence of living cells, (2) invasion of the site by native osteoblasts, and (3) the transformation of host cells into osteoblasts. Bone grafts may be autologous, cadaver donor, or an artificial calcified matrix. Coral has been used successfully as a bone substitute. Autologous bone is always preferable. For significant bony defects, microvascular osseous-free flaps can provide the best functional result with the most bone.

The surgeon can best avoid the pitfalls of nonunion, malunion, and fibrous union and enhance the process of facial bone wound healing by relying on three basic principles: (1) creating optimum bone to bone contact (reduction), (2) rigid fixation, and (3) immobilization for an adequate period to allow healing. Rigid fixation with plates and screws, with or without bone grafts for large defects, provides a treatment that adequately addresses all three principles.

Tympanic Membrane

The tympanic membrane (TM) possesses a powerful ability to repair itself due to the unusual features of its epithelial structure and migratory capacities. It is composed of three histological layers. The outer layer consists of keratinizing stratum corneum. The middle section is a thin layer of connective tissue. The innermost layer consists of a mucosal layer of cuboidal surface epithelium. The concept of epithelial generation entails the migration of epithelium around the TM in a centripetal fashion from anterior to posterior, then laterally along the canal wall. The TM heals in a process similar to the cutaneous process described previously. It differs from skin with respect to wound healing during the proliferative phase. In the skin, the fibrous connective tissue layer precedes the migration of the epithelium across the wound. Conversely, in the TM, the epidermal layer plays the critical initial role in migration, whereas the fibrous layer is usually the last layer to cross the defect and in many cases does not, resulting in a dimeric membrane, commonly referred to by the misnomer "monomer."

The majority of traumatic TM perforations will heal spontaneously without intervention. Several factors inhibit or delay closure of the TM, resulting in a chronic perforation. A chronic perforation is caused by the growth of squamous epithelium over the edges of a perforation to meet the mucosal layer. This epithelialized wound edge forms a fistula between the middle ear and the canal. Chronic perforation can occur due to severe acute infection, chronic infection, large defects secondary to trauma, presence of a burn, concurrent steroid administration, old age, and poor nutritional status.

Fetal Healing

The key to solving the problem of scar formation lies in an increased understanding of fetal healing. Although the field is in its infancy, successful treatment of such lethal and near-lethal defects as hydronephrosis, hydrocephalus, lung hypoplasia, diaphragmatic hernia, myelomeningocele, sacrococcygeal teratoma, laryngeal atresia, and stenosis is now possible and becoming more frequently performed and widely available. Risks to the mother and fetus still limit fetal surgery for conditions such as cleft lip and palate, but they will certainly be one of the next applications.

Specific aspects of fetal healing believed to contribute to the scarless regeneration that has been observed include faster cell proliferation, lower oxygen tension, the presence of a fluid environment (amniotic fluid), sterile conditions, decreased inflammatory response, a faster and more organized matrix deposition, and lower concentrations of growth factors TGF-α, bFGF, higher levels of insulin-like growth factor and hyaluronic acid stimulating factor, and reduced levels of neoangiogenesis following injury. In general, studies indicate that it is the fetal cells themselves that provide the result, rather than the environment. As gestation progresses, regeneration capacity decreases, and scarless healing ends during the second trimester. This correlates with an increasing inflammatory response to injury.

The fetus does have the capacity to form scar tissue at an early age, with intestinal adhesions found following intrauterine diaphragmatic hernia repair, while at the same time, the thoracic incision heals with no scar. Organ differences in healing, age-related differences, and the variety of distinctive biochemical and cellular activities that characterize fetal healing bring new insights into tissue repair and regeneration and provide tantalizing insight into all healing, but they have not provided the key to successfully reducing scar formation after birth.

PERIOPERATIVE PREPARATION FOR SUCCESSFUL WOUND HEALING

The surgical outcome is predicated on the skill of the surgeons and anesthesiologists, the nature of the disease being treated, and the condition of the patient. Careful attention to preparation for surgery will optimize the likelihood of a satisfactory result. Pretreatment of systemic conditions will significantly reduce the incidence of postoperative complications, especially wound healing problems and infections. Conditions including hypertension, malnutrition, pulmonary insufficiency, diabetes mellitus, hypothyroidism, smoking cessation, and existing skin, sinus, or urinary tract infections should be treated. Treatment with vitamins A and C in those patients on chronic steroids should be initiated before surgery. Assuring adequate blood volume, maintaining body temperature (avoid a cold operating room at the time of induction) to minimize thermoregulatory vasoconstriction, and treating pain and anxiety will facilitate tolerance of anesthesia and reduce complications.

During surgical procedures attention should focus not just on the procedure but also on several other aspects of patient management. These include maintaining antibiotic levels (if necessary) and body temperature, keeping all tissues in the surgical field moist, avoiding undo pressure on any parts in the surgical field or on the patient on the table, irrigating the wound with antibiotic solutions if contamination has occurred, using cautery with care to avoid leaving charred or necrotic tissue in the wound, and considering a delayed wound closure if the field is heavily contaminated. Discussions with the anesthesiologist should assure that fluid replacement keeps abreast of fluid losses, that systemic mean arterial blood pressure remains stable, and that urine output is adequate.

Postoperative attention to details, such as keeping the patient warm and pain free, optimizing oxygen delivery, maintaining antibiotic levels, instituting nutrition quickly, using parenteral or enteral nutritional supplements if needed, and continuing treatment of coexisting conditions like hypertension and diabetes, will all improve the quality of outcome from the surgical treatment.

Suggested Readings

Achaur, BM, Erickson E. Principles and techniques. In: Plastic Surgery: Indications, Operations and Outcomes. St. Louis: Mosby; 2000

Granick MS, Long CD, Ramasastry SS. Wound healing state of the art. Clinics Plas Surg 1998;25:3

Hom DB, Szachowicz. Wound healing for the otolaryngology's–head and neck surgeon. Otolaryngol Clin NA 1995;28(10):5

Hunt TK. The role of the macrophage in wound healing. Surg Forum 1976;27(62):16–18

Manjo G. The Healing Hand: Man and Wound in the Ancient World. Cambridge, MA: Harvard University Press; 1975

Niinikoshi JH. Clinical hyperbaric oxygen therapy: wound perfusion and transcutaneous oximetry. World J Surg 2004;28(3):307–311

Wattel F, Matthieu D, Coget JM, et al. Hyperbaric oxygen in the management of chronic vascular wounds. Angiology 1990; 41(1):59–65

Witte MB, Barbul A. General principles of wound healing. Surg Clin North Am 1997 June; 77(3):509–528

Self-Test Questions

For each question select the correct answer for the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. Wound healing is initiated by cellular and biochemical reactions. Critical events involved in the normal wound healing process include all but one of the following:
   A. The extrinsic pathway is initiated by tissue factor, binding to factors VII and VIIa. Factor XIII (fibrin-stabilizing factor) initiates fibrin clot development.
   B. The intrinsic coagulation path is activated via factor XII, when blood is exposed to foreign surfaces.
   C. Thrombin clot within the wound provides the provisional matrix for the extracellular matrix (ECM) that becomes the scaffolding for subsequent cellular migration, adhesion, and proliferation.
2. The first cell type that migrates into the wound following injury is
   A. Macrophage
   B. Platelet
   C. Neutrophil
   D. Fibroblast
3. The cell responsible for the production of glycosaminoglycans, hyaluronic acid, chondroitin-4 sulfate, dermatan sulfate, and heparin sulfate substrates for ground substance is
   A. Neutrophil
   B. Macrophage
   C. Platelet
   D. Fibroblast
4. Incisions made parallel to the relaxed lines of skin tension will
   A. Most likely form keloids
   B. Widen significantly in contrast to those made at a right angle
   C. Never hypertrophy
   D. Result in satisfactory healing

Chapter 3
BASIC PRINCIPLES OF ALLERGIC DISEASES

David Rosenstreich, Ashok Vaghjimal, and Golda Hudes

Allergic diseases (allergic rhinitis, asthma, urticaria, and atopic dermatitis) are very common, affecting 20 to 25% of the U.S. population. They are the sixth leading cause of chronic disease in the United States, with a cost of more than $10 billion each year.

A 1993 survey revealed that allergic rhinitis alone affects 39.9 million persons in the United States. This results in ~811,000 missed workdays, 824,000 missed school days, and 4,230,000 reduced activity days. A high prevalence of allergic rhinitis has also been reported in England, Australia, and Sweden.

This chapter will discuss current understanding of the pathophysiology, diagnostic approaches, and therapy of allergic diseases, with an emphasis on allergic rhinitis.

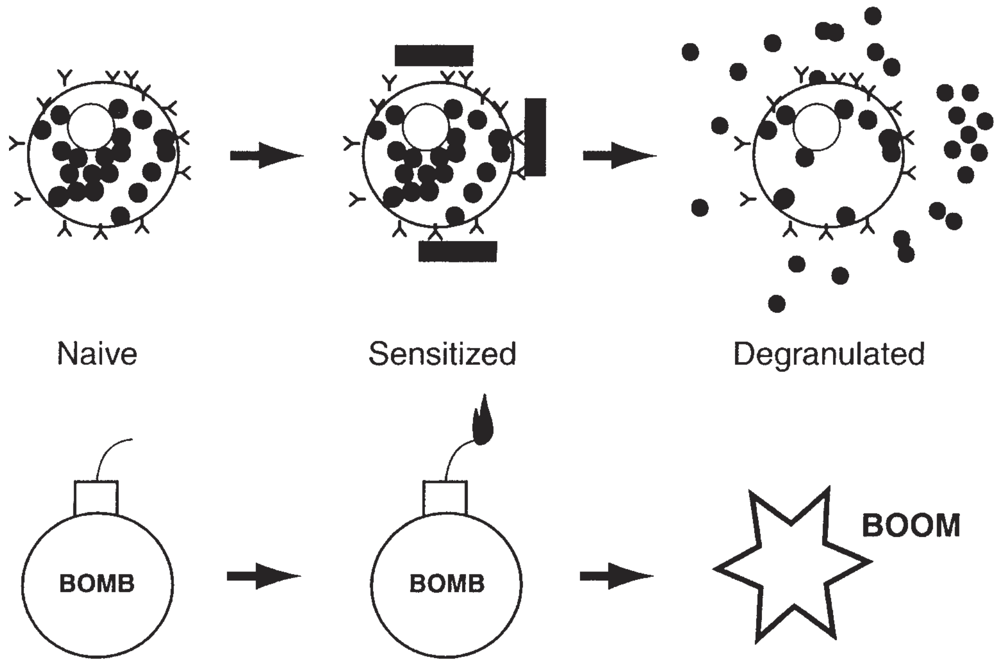


Figure 3–1 *Allergic sensitization and degranulation. The process of sensitization and degranulation in mast cells is analogous to the detonation of a bomb. Initial binding of specific IgE to the naive mast cell surface "primes" the cell for activity. Subsequent binding of allergen to the mast cell is akin to lighting the fuse of the bomb. Intracellular biochemical events lead to the ultimate "explosion," a cellular degranulation leading to mediator release. (From Allergic Diseases: Diagnosis and Treatment, edited by Phil Lieberman and John A. Anderson, Humana Press, Totowa, New Jersey, 1997. Reprinted with permission.)*

PATHOPHYSIOLOGY OF ALLERGIC DISEASES

The tendency to develop allergy is inherited and is termed *atopy*. The process by which one initially becomes allergic to a substance begins with sensitization, which is a two-step process (**Fig. 3–1**). During the initial stage of sensitization, the allergen is processed by antigen-presenting cells (macrophages, dendritic cells) and presented to T lymphocytes. T lymphocyte–derived interleukins (ILs), especially IL-4 and IL-13, induce immunoglobulin E (IgE) production by B cells and plasma cells in the presence of the allergen. This newly formed IgE adheres to tissue mast cells or to circulating blood basophils. This process rarely produces any symptoms.

The second step in this process involves the reexposure of a sensitized individual to the allergen. Cross-linking of at least two adjacent IgE molecules by the allergen induces mast cell/basophil release of histamine and other proinflammatory mediators within minutes. These mediators cause the typical allergic symptoms, which may range from mild (itching, sneezing) to severe (anaphylaxis), resulting in death.

This immediate reaction to an allergen (the early-phase reaction) cannot fully explain allergic symptoms, which may last for hours to days. This mechanism also cannot explain the therapeutic efficacy of corticosteroids, which lack the ability to inhibit immediate responses to an allergen. A large body of research indicates that the immediate allergic reaction is usually followed 2 to 8 hours later by a late-phase reaction, which produces the chronic or prolonged symptoms. This reaction can develop in the nose, respiratory tract, skin, and other anatomical locations, and is marked by infiltration of many types of cells, primarily eosinophils, but also basophils, mononuclear cells, and neutrophils. All these cells contribute to the late-phase reaction by producing additional proinflammatory mediators and cytokines.

CELLULAR COMPONENTS OF THE ALLERGIC RESPONSE

Mast Cells and Basophils

The major effector cells of allergic reaction are mast cells and basophils. Both cell types are derived from hemopoietic progenitor cells in the bone marrow and are the major source of histamine and other allergic mediators. Morphologically, human basophils have a segmented nucleus with marked condensation of nuclear chromatin and contain round or oval cytoplasmic granules.

In contrast, mast cells typically appear as either round or elongated cells with a nonsegmented nucleus.

Basophils mature in the bone marrow and then circulate in the blood. Mast cells, however, leave the circulation and mature and reside predominantly in connective tissues. There are two types of mast cells in humans: mast cells containing the enzyme tryptase (MCTs) and mast cells containing both tryptase and other enzymes, including chymase, mast cell carboxypeptidase, and cathepsin G (MCTCs). MCTs reside mainly in the lungs and intestinal mucosa, and MCTCs in the skin and intestinal submucosa. It is thought that cytokine differences in the microenvironment of the tissues are responsible for this cellular differentiation. Precursor cell differentiation to MCT (mucosal) mast cells is the result of mature cytokines such as stem cell factor (SCF). It is not clear what causes a mast cell to develop as either a MCT (mucosal mast cell) or MCTC (connective tissue mast cell). However, *in vitro* studies suggest that T lymphocycte associated activity appears to be important in the development of MCT cells, whereas SCF and other fibroblast cytokines. Fibroblast cytokines are important in the development of both types of mast cells.

In addition to containing mediator-rich granules, the most important distinguishing feature of mast cells and basophils is their ability to tightly bind the allergy-inducing antibody IgE. Mast cells and basophils both possess a high-affinity receptor for IgE, Fc(RI), and are the principal cells in the body with these receptors. Once bound to Fc(RI), IgE can remain on the cell surface for many months. Other cells, such as B cells and activated T cells, possess a low-affinity IgE receptor, Fc(RII), which has been identified as the CD 23 molecule. However, the exact role of this low-affinity IgE receptor in allergy is not known.

The most important cellular event that underlies expression of basophil or mast cell function is degranulation, which occurs by cross-linking of Fc(RI) receptors. Bridging of these receptors through bound IgE molecules on the mast cell or basophil causes several intracellular biochemical events catalyzed by phospholipase C and phospholipase A_2. These events culminate in mast cell/basophil degranulation and activation, with release of various stored or newly formed mediators such as histamine, leukotrienes, and prostaglandin D_2. Monovalent antigens will not trigger mast cell or basophil activation.

In addition to IgE and specific antigen, a variety of substances can directly elicit release of mediators from mast cells and basophils. A few examples of these include physical stimuli such as cold, exercise, and sunlight; drugs such as vancomycin and opiates; and chemical compounds such as radiological contrast dyes.

Eosinophils

Eosinophils play a major role in chronic allergic reactions as effector cells. They are derived from bone marrow and are similar in size to neutrophils but have bilobed nuclei and distinctive cytoplasmic granules. They are most abundant in tissues that have a mucosal epithelial interface with the environment, such as the respiratory and gastrointestinal tracts. The major mediators stored in eosinophil cytoplasmic granules include major basic protein, eosinophil cationic protein, eosinophil-derived neurotoxin, eosinophil peroxidase, lysosomal hydrolase, and lysophospholipase. The latter enzyme forms the Charcot-Leyden crystals that are found in the sputum of asthmatics and which are evidence of prior eosinophil degranulation.

The development of eosinophils is promoted by at least three cytokines. Granulocyte-macrophage colony–stimulating factor (GM-CSF), IL-3, and IL-5, have the most cell-specific effects on eosinophil differentiation and production. IL-5 also rapidly increases eosinophil release from the bone marrow into the circulation.

Human eosinophils express receptors for IgG, IgA, and IgE. Receptors for complement components and several cytokines have also been found on eosinophils. Though the exact protective function of eosinophils is still unclear, they are thought to play an important role in defense against the helminthic parasites.

Eosinophils are major contributors to the pathophysiology of many allergic disorders, including allergic rhinitis. Many studies have demonstrated the influx of eosinophils during late-phase allergic reactions. For instance, there is a significant increase in the total number of eosinophils and the number of activated eosinophils in the epithelium and lamina propria in the nasal biopsy specimens of patients with grass pollen allergy after an experimental, out-of-season, 2-week allergen exposure. However, the majority of studies have failed to document a correlation between symptom scores, pollen counts, released mediators, and the total number of eosinophils or number of activated eosinophils, which suggests that other cells in the nasal mucosa along with their mediators are also involved in producing allergic symptoms.

T Lymphocytes

T lymphocytes are important regulators of allergic reactions. As described previously, they are essential for the production of IgE and for the production of cytokines that activate and differentiate mast cells and eosinophils. During ontogeny in the thymus, T cells eventually develop into either CD4+ or CD8+ cells. CD4+ cells, which represent approximately two thirds of peripheral blood

T lymphocytes, recognize antigens bound to class II human leukocyte antigen (HLA) molecules. CD8+ lymphocytes, which represent approximately one third of peripheral blood T cells, recognize antigens bound to class I HLA molecules.

CD4+ T lymphocytes are often associated with helper function, and CD8+ cells are believed to have suppressor and cytotoxic properties, although these correlations are not absolute.

Five to 10% of peripheral blood lymphocytes are CD4–CD8. It is thought that they may play a role in the recognition of glycolipids.

Helper T lymphocytes are divided into Th1 and Th2 cells based on their cytokine production patterns. Th2 cells primarily produce IL-4, IL-5, IL-9, IL-10, and IL-13. They play an important role in the pathogenesis and promotion of allergic diseases and humoral immunity. IL-4 induces the differentiation of Th2 cells from naive T helper cells (To) and inhibits the differentiation of Th1 lymphocytes.

Th1 cells, in contrast, act primarily to suppress allergic mechanisms and reactions. They induce and promote cell-mediated inflammation by secreting IL-2, tumor necrosis factor β (TNF-β), and interferon γ. The latter cytokine has an inhibitory effect on the differentiation of Th2 cells and IgE production. IL-12 from macrophages promotes the differentiation of Th1 cells from the To population.

The discovery of the Th1 and Th2 cells has led to an interesting explanation of increasing prevalence of allergic diseases in Western countries. According to this theory, exposure to infections early in life tends to favor Th1 immune responses and suppresses Th2 responses. Improved immunization, better control of infectious diseases, and widespread use of antibiotics in Western countries may encourage Th2 responses and leads to a higher prevalence of allergic diseases. A recent observation that children who responded well to bacille Calmette-Guérin (BCG) vaccination were less likely to develop IgE response later in life supports this view. Conversely, there is a decreased prevalence of allergy and asthma in younger siblings presumably because they are exposed to more childhood infections at an earlier age.

B Lymphocytes

The major contribution of B lymphocytes to allergic reactions is through their production of IgE. B cells mature in the bone marrow and differentiate into plasma cells, which are responsible for most of the immunoglobulin production in the body, including IgE. This production is primarily T cell mediated. Thus activation of Th2 cells with IL-4 elaboration induces a B cell switch from IgM production to IgE and IgG4 production, and transforming growth factor β (TGF-β) in combination with IL-10 induces a switch from IgM production to IgA1 and IgA2 production.

Most IgE-producing B cells and plasma cells are located in tissues where allergic reactions occur, and most IgE is produced locally. This is in contrast to the majority of the other immunoglobulins, which are produced in secondary lymphoid organs, the spleen and lymph nodes.

Allergic Antibodies

Immunoglobulin E plays a central role in allergic disease. The ability to produce large amounts of this antibody distinguishes atopic from nonatopic individuals because the latter individuals usually produce other classes of antibodies (IgG, IgM) when exposed to an antigen. Studies have shown that total serum IgE concentrations tend to be higher in allergic adults and children compared with nonallergic individuals, but the diagnostic value of this test is limited.

The production of IgE is mainly under the control of two cytokines, IL-4 and IL-13, which are produced by Th2 cells. Indeed, IL-4 is such a necessary signal for IgE production that mice genetically engineered to be devoid of IL-4 (IL-4 knockout mice) are unable to synthesize IgE. Although several studies have demonstrated that only IL-4 and IL-13 are capable of inducing IgE synthesis, IgE production can be further enhanced or inhibited by a number of cytokines. Thus IL-2, IL-5, IL-6, IL-9, and TNF-*a* synergize with IL-4 and IL-13 to enhance IgE production.

Other cytokines, produced primarily by Th1 cells, tend to downregulate IgE synthesis. The IgE isotype switch is inhibited by interferon γ (IFN-γ) and TGF-β. IL-12 and IL-18 inhibit the differentiation of IL-4-producing T cells. IL-8 also inhibits IgE production by purified B cells, but the molecular targets of this process remain unclear. Interestingly, IFN-α, IFN-γ, and IL-12 inhibit IgE synthesis only when tested in a T cell–dependent system (e.g., IL-12 does not suppress IgE synthesis by purified B cells), and IL-10 suppresses IgE production only in the presence of monocytes (i.e., it fails to inhibit IL-4-induced IgE synthesis by highly purified B cells in the absence of monocytes).

On a molecular level the initiation and inhibition of IgE synthesis involve a very complex cascade of metabolic pathways that ultimately inhibit or initiate gene transcription. This process uses different groups of enzymes. One of them, signal transducer and activator of transcription 6 (STAT6), promotes IL-4 signaling and the development of Th2 cells.

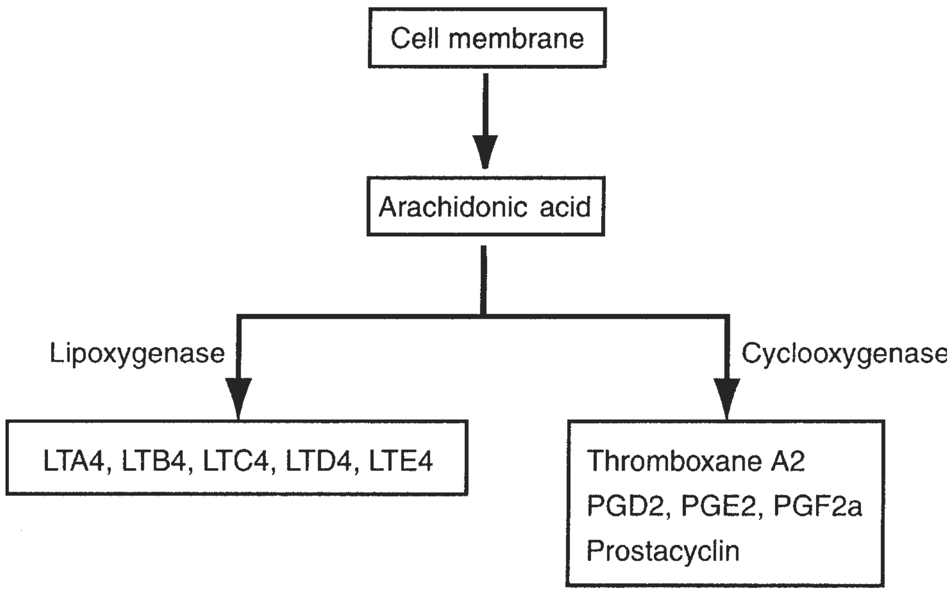


Figure 3–2 *Arachidonic acid metabolic pathways. Arachidonic acid, released as a result of phospholipase action on the cellular membrane, is broken down by two distinct biochemical pathways. The lipoxygenase pathway results in formation of the leukotrienes, whereas the cyclooxygenase pathway generates prostacyclin, thromboxane, and the prostaglandins. (From Allergic Diseases: Diagnosis and Treatment, edited by Phil Lieberman and John A. Anderson, Humana Press, Totowa, New Jersey. Reprinted with permission.)*

ALLERGIC MEDIATORS

Histamine

Histamine is the major mediator of acute and early-phase allergic reactions. It is contained in a preformed state within the granules of mast cells and basophils and is formed from the amino acid histidine. Its activities include dilatation and increased permeability of blood vessels, constriction of bronchial airways, stimulation of nerve endings, and secretion of mucus in airways. In extreme cases, histamine-induced intravascular fluid shifts can lead to hypotension and shock.

Leukotrienes

Leukotrienes (LTs) are the major mediators of the late phase and chronic allergic reactions. They are the products of arachidonic acid metabolism and are formed by the action of a lipoxygenase enzymes (**Fig. 3–2**). They are produced by many cell types, including mast cells, basophils, macrophages, and eosinophils, but on a per cell basis, eosinophils produce more leukotrienes than any other cell type. On a molar basis, leukotrienes are 100 times more potent bronchoconstrictors than histamine. In addition, unlike histamine, whose tissue effects are short-lived, leukotriene-induced tissue activation lasts a relatively long time. Because of these properties, leukotrienes are thought to be the major mediators of chronic allergic diseases such as asthma. They induce a variety of effects, including constriction of the bronchial airways and increasing the permeability of small blood vessels through stimulation of specific receptors.

Two major classes of these receptors, CysLT1 (for cystinyl leukotriene) and BLT (for LTB4), have been described. Leukotrienes transduce signals via the CysLT1 receptor (previously known as LTD4 receptor). In isolated human airway smooth muscle, LTC4 and LTD4 are of equal potency as contractile agonists at the CysLT1 receptor, whereas LTE4 is less potent.

LTB4 is a natural ligand for the BLT. When activated, the BLT receptor transduces signals that result in chemotaxis and cellular activation.

Prostaglandin D

Prostaglandin D (PGD) is also formed in human mast cells from the metabolism of arachidonic acid, but through the cyclooxygenase rather than the lipoxygenase pathway. It causes constriction of bronchial airways. Elevated amounts of PGD are found both in nasal lavage fluid and in bronchoalveolar lavage after allergen challenge. However, PGD is only elevated during the early response and does not play a role in maintaining the late response or chronic allergic inflammation.

Cytokines

Cytokines are a diverse group of glycoproteins that are synthesized and secreted by many cell types in response to their activation or injury. As discussed previously, IL-4 and IL-13 cause the upregulation of the IgE response, and IL-12 together with IL-18 (via IFN-γ production) downregulates IgE synthesis. In addition, numerous cytokines play important roles in developing and maintaining allergic responses. For instance, TNF- and IL-4 cause the enhancement of basophil recruitment. IL-5, IL-3, and GM-CSF promote eosinophil differentiation, activation, and survival. Several chemokines, small specific chemotactic cytokines such as RANTES, and eotaxin are highly selective for eosinophils and play a major role in the recruitment of eosinophils to sites of allergic inflammation. Another chemokine, IL-8, is a major chemoattractant for neutrophils and plays an important role in chronic allergic inflammation.

Neutral Proteases

Neutral proteases are present in the secretory granules of mast cells and basophils. They are released after mast cells are activated and can be measured in the serum after systemic allergic reactions. They help in the degradation of proteins by attacking peptide bonds. As described previously, the two major proteases generated by mast cells and basophils are tryptase and chymase. Tryptase is located in both the MCTC cells and MTC subtypes of mast cells, and chymase is located only in the MCTC cells.

Proteoglycans

The major proteoglycans in mast cells are heparin and chondroitin sulfate A. Their role in allergic reactions is the regulation of mediator release from the granules. They bind both histamine and proteolytic enzymes and are believed to stabilize them until granulation occurs. Although they are released and can be detected after allergic reactions, they do not seem to have any effector functions. The activity of various mediators is summarized in **Table 3–1**.

Table 3–1 Cytokines involved in allergic inflammation

Cytokines	Source	Activity
IL-4, IL-13	Th2 lymphocytes, mast cells, basophils	IgE isotype switch, Th2 lymphocyte proliferation
IL-5	Th2 lymphocytes, eosinophils	Eosinophil activation and survival
Eotaxin	Th2 lymphocytes	Eosinophil chemotaxis
Interferon-gamma	Th1 lymphocytes	Inhibits IL-4 action and IgE production
IL-10	Th2 lymphocytes, mast cells, eosinophils, B lymphocytes	Downregulates IFN-γ, II-2, and TNF-β production
Major basic protein	Eosinophils	Allergic protein inflammation, epithelial cell destruction
IL-12	Macrophages	Inhibition of IgE synthesis via IFN-γ production

Ig, immunoglobulin; IFN, interferon; IL, interleukin; Th, T lymphocyte; TNF, tumor necrosis factor.

ALLERGENS

Airborne allergens are the most important inducers of allergic rhinitis in both children and adults. **Table 3–2** presents the most common allergens responsible for the development of allergic symptoms. Sensitization with indoor and outdoor aeroallergens is the most common cause of seasonal and perennial allergic rhinitis.

Dust Mites

House dust mites were first described in 1927 and clearly identified as a cause of asthma in 1967. They are the most important indoor allergens in many areas of the world. Ideal growing conditions for mites are a relative humidity over 50%, temperatures around 70°F, and the presence of shed human skin, which is their major food source. Almost all of the major mite allergens (Der P and Der F) are contained in mite fecal pellets. Because of these growth requirements, bedding has the highest concentration of mite allergens in the home. It has been estimated that there are approximately 250,000 mite fecal pellets in every ounce of dust in a mattress. Most individuals in the developed world therefore spend about 8 hours every night of their entire lives breathing in mite feces. It is therefore not surprising that dust allergy is so prevalent. The majority of patients with perennial allergic rhinitis throughout the world are allergic to dust mites.

Animal Allergens

Although the dander of any mammalian pet can trigger allergic symptoms, cat allergen is probably most common and best studied. Unlike mite allergens, the major cat allergen, Fel d 1, is carried on small particles (~5μc) and can be detected in the air, on wall surfaces, and on clothing. Other animals, such as birds, cows, and horses, have been reported as sources of allergens for atopic patients.

TABLE 3–2 ALLERGENS RESPONSIBLE FOR PRECIPITATING ALLERGIC SYMPTOMS

- I. Environmental allergens
  - 1. Indoor allergens
    - a. Dust mites
    - b. Furry or feathered pets
    - c. Feathers/down
    - d. Cockroaches
    - e. Rodents
  - 2. Outdoor allergens
    - a. Pollen
    - b. Mold spores
- II. Food allergens
  - 1. Most common
    - a. Nuts and peanuts
    - b. Shellfish
    - c. Eggs
    - d. Cow's milk
    - e. Fish
  - 2. Less common
    - a. Corn
    - b. Wheat and other grains
    - c. Soybean and other legumes
    - d. Fruits such as strawberry, kiwi, and banana
- III. Drug allergens
  - 1. Haptens
    - a. Penicillin and other antibiotics
    - b. Sulfonamides
    - c. NSAIDs
  - 2. Proteins and other high-molecular-weight drugs
    - a. Insulin
    - b. Psyllium
- IV. Occupational allergens
  - 1. Reactive chemicals
    - a. Anhydrides (phthalic, timelitic, tetrachlorphthalic)
    - b. Diisocyanates (polyurethane)
    - c. Wood dusts
  - 2. Proteins and other high-molecular-weight allergens
    - a. Laboratory animals
    - b. Enzymes (subtilsin, trypsin)
    - c. Latex
    - d. Psyllium

NSAIDs, nonsteroidal anti-inflammatory drugs.

COCKROACHES

Cockroaches are a very important cause of respiratory allergy among impoverished inner-city populations. Cockroach allergens are found on particles of similar size to mite allergens, and their highest concentration is detected in the kitchen. However, recent data indicate that the bedroom is the most important room in the home in terms of cockroach allergen exposure and resultant sensitization.

MOLDS (FUNGI)

Molds are highly ubiquitous and are common causes of allergic rhinitis. *Cladosporium* and *Alternaria* species are the best recognized fungal offenders. However, other types, such as *Penicillium, Rhizopus,* and *Aspergillus,* are also frequently implicated as allergens. Although fungi can be detected in the air throughout the year, they are increased during warm, humid seasons. Their seasonal prevalence in the Midwest region of the United States is shown in **Fig. 3–3**.

Although fungus exposure usually comes from outdoor sources, mold also can be a source of perennial indoor exposure, often growing on shower curtains, in damp basements, and on indoor plants.

POLLENS

Only a minority of plants produce windborne pollen, and even fewer cause nasal allergy. There is a defined sequence of pollen shedding that differs in various climatic regions of the world **(Fig. 3–3)**. In the northern United States, tree pollens of familiar deciduous trees such as oak, maple, elm, and birch appear early in the spring, before the leaves unfold. Grass pollen follows that of trees. Finally, there is a late-summer peak of weeds and ragweed pollen. In warmer areas of the United States, such as the Southeast and Southern California, the grass season may last 6 months or more. In contrast, some areas, such as Southern California and parts of Europe, are devoid of ragweed and do not have a late-summer, "hay fever" season.

INGESTANTS

Although food and drug allergies are common causes of allergic skin, gastrointestinal, and systemic reactions, isolated respiratory reactions due to food and drug allergy are less common. However, there are possible cross-reactions between shrimp and the inhalant or ingested allergens in insect materials such as cockroaches, grasshoppers, and fruit flies. Patients with pollen allergy can develop oral itching on ingestion of cross-reacting fruits. This is termed the *oral allergy syndrome* and is seen with fruits such as cantaloupe in ragweed-sensitive patients or apple in patients allergic to birch pollen.

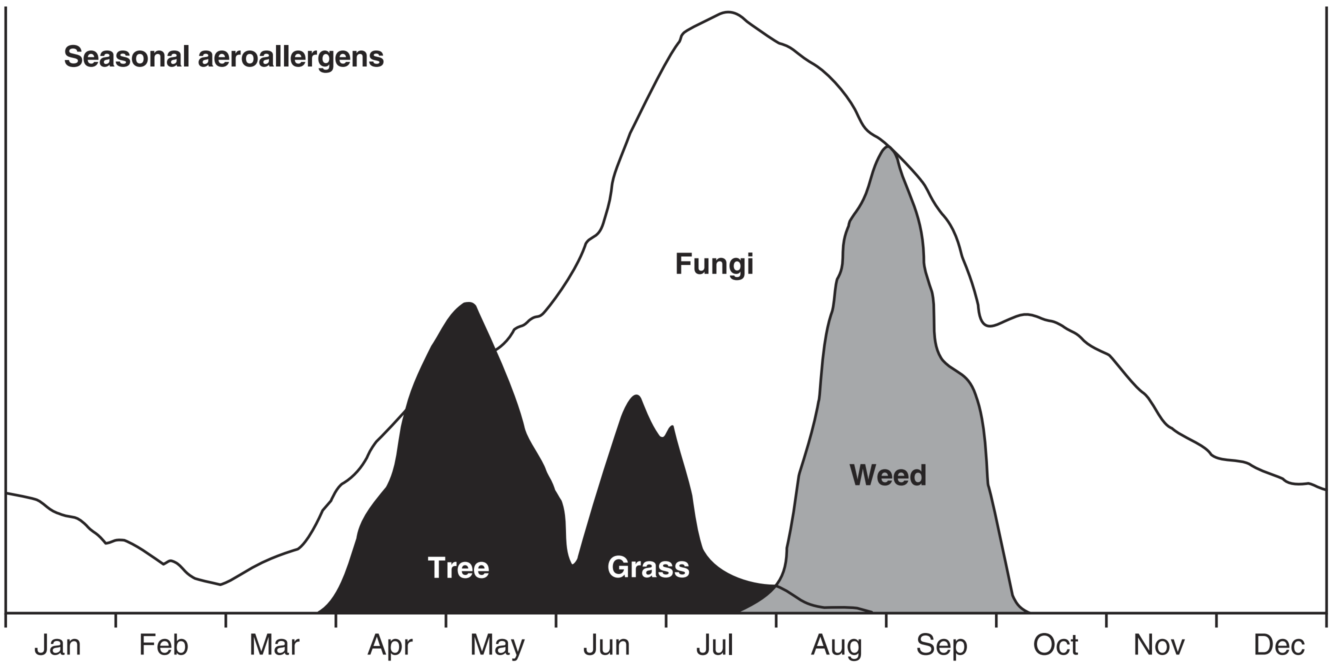


Figure 3–3 *Prevalence patterns for Midwestern pollens and fungi outdoors. Frost occurs from the middle of September through March, and snow occurs from November through February. (From Robert M. Naclerio & William Solomon; Rhinitis and Inhalant Allergens, JAMA, Vol 278, No22, 1997, 1844. Reprinted with permission)*

Occupational Allergens

IgE-mediated reactions to latex have drawn special attention recently because of systemic and respiratory reactions in exposed patients and health care workers. Occupational IgE-mediated reactions can be caused by laboratory animal products, grain (in bakers and agricultural workers), coffee beans, wood dust such as mahogany and western red cedar, chemicals such as platinum dust, and psyllium. Occupational rhinitis may precede the development of occupational asthma.

CLINICAL PRESENTATION

Allergic rhinitis often presents with repetitive sneezing, rhinorrhea, nasal congestion, and itching of the mucous membranes. In severe cases, ear fullness and popping, malaise, and fatigue may be present. The symptoms of nonallergic rhinitis can be similar.

Chronic allergic rhinosinusitis is often a little more difficult to diagnose on clinical history alone. Usual complaints are dull pressure in the face or head, nasal congestion, and postnasal drip, but patients may also report cough, hypo-osmia, sore throat, fetid breath, and malaise. Children frequently have an earache due to coexisting otitis media. A good history and physical examination play a crucial role in diagnosing allergic rhinitis. Specific timing of symptoms suggests "hay fever" or seasonal allergic rhinitis. One should remember that mold spores can cause seasonal symptoms as well. The link of perennial symptoms with a specific allergen is less obvious and may require special evaluation including in vivo or in vitro allergen testing.

Physical examination is important, but it lacks specificity. Classical clinical signs such as allergic shiners, Dennie-Morgan infraorbital creases, and the "allergic salute" are frequent in children, but they are rarely seen in adults. Examination of the nasal mucosa can reveal pale, swollen nasal turbinates, but some patients may have an erythematous mucosa. After shrinking of swollen mucosa with topical decongestants, polyps can often be readily appreciated.

Allergic rhinitis often coexists with allergic conjunctivitis. The role of allergy in otitis media is controversial, although eustachian tube dysfunction often occurs when rhinitis is prominent.

Other relevant organ system findings include examination of the chest for signs of asthma and the skin for concomitant eczema and atopic dermatitis.

TABLE 3–3 EVALUATION OF AN ALLERGIC PATIENT

History
  a. Recent illness
  b. Family history
  c. Occupational history
Physical exam
  Skin testing
  a. Percutaneous (prick and puncture)
  b. Intradermal
Laboratory exam
  a. Eosinphil count (peripheral blood and nasal lavage)
  b. Radioallergosorbent test

DIAGNOSTIC APPROACHES AND ALLERGEN TESTING

Table 3–3 presents the diagnostic approaches to a patient with allergic rhinitis. A careful history and physical exam are the most important steps in the diagnosis of allergic rhinitis. The demonstration of specific IgE antibodies is often useful in diagnosing and better characterizing allergic rhinitis.

Allergy testing is done to identify trigger allergens, reinforce environmental control instructions, and, in some cases, establish a need for direct immunotherapy. Testing is commonly done by one of two ways, skin testing or radioallergosorbent test (RAST).

SKIN TESTING

Skin testing, which requires trained personnel, is usually done in experienced medical facilities because of issues of interpretation and techniques. However, testing is usually safe, and the incidence of adverse reactions is very low. The number of tests depends on the range of local allergens and a particular patient's exposure. If initial prick test results are negative, the more sensitive intradermal method is usually tried. Overall, skin tests show greater sensitivity than serum assays. They are quick, safe, and cost-effective.

RADIOALLERGOSORBENT TEST

Radioallergosorbent test is an in vitro test assessing serum levels of specific IgE antibodies. It should be utilized in patients who fear skin testing or have dermatographism or extensive skin disease, and in those who must take medications that interfere with skin testing (e.g., antihistamines, certain antidepressants, extensive topical corticosteroids, and possibly beta-blockers). Reliable RAST is now available from most commercial laboratories. Tests for individual specific allergens can be ordered or can be ordered as panels. The components of panels vary between laboratories.

THERAPY

Effective treatment of allergic rhinitis involves a combination of environmental control and pharmacotherapy. When symptoms persist despite these measures, allergen immunotherapy should be considered.

ENVIRONMENTAL CONTROL

Environmental control is the most useful method of reducing allergy symptoms in most patients and should be instituted in everyone. All surroundings, particularly the bedroom, of the patient should be as free as possible from known and potential asthma triggers. This is the most effective and important step in the treatment of any allergic disease.

PHARMACOTHERAPY

Oral antihistamines are effective in reducing symptoms of itching, sneezing, and rhinorrhea and are the first-line therapy for allergic rhinitis. However, they have little effect on nasal congestion. Therefore, there are a number of antihistamines, both first and second generation, available combined with a decongestant.

Intranasal antihistamines such as Astelin are effective and in contrast with oral preparations may reduce nasal congestion. They are appropriate for use as first-line treatment of allergic rhinitis.

Those with prolonged allergic symptoms should add intranasal corticosteroids. They are the most effective medication class in controlling symptoms of allergic rhinitis. All the newer topically acting agents are generally safe and are not associated with significant systemic side effects in adults. Nasal irritation and bleeding may occur, and very rare cases of nasal septum perforation have been reported. Most reported cases occurred with older preparations

A short course of oral corticosteroids is sometimes used to treat very severe or intractable nasal symptoms or nasal polyps. However, their repeated or prolonged use in allergic rhinitis should be discouraged.

Intranasal cromolyn sodium is effective in some patients and is associated with minimal side effects. It is most useful as a prophylactic agent (i.e., to prevent symptoms when an allergenic exposure is anticipated). Intranasal anticholinergics (Ipatropium bromide) may also effectively reduce rhinorrhea but usually have no effect on other nasal symptoms.

There is some preliminary evidence that leukotriene antagonists may be useful in allergic rhinitis, but the exact role of this class of drugs remains to be established.

For patients with persistent eye complaints, drops containing cromolyn, antihistamines, or nonsteroidal anti-inflammatory drugs complement oral antihistamines. Intraocular steroids are less often used. They are potentially harmful and require close ophthalmological follow-up.

Allergen Immunotherapy

Allergen immunotherapy may be highly effective in controlling symptoms of allergic rhinitis. A decision to start immunotherapy is based on the physician's cost–risk benefits analysis and the patient's preference. This mode of treatment is usually initiated for several reasons, including severe seasonal or perennial rhinoconjunctivitis in which optimal allergen avoidance and medication have not been sufficiently effective in controlling symptoms; rhinitis caused by allergen that cannot be avoided (e.g., occupational exposure to laboratory animals, pets at home); the use of daily pharmacotherapy, such as systemic corticosteroids for prolonged periods; and intolerable side effects from medications.

Recent studies suggest that immunotherapy may prevent further development of asthma in patients with allergic rhinitis. Allergen immunotherapy can be stopped in most patients after 3 to 5 years with persistence of therapeutic effects.

Investigational Treatments

As IgE regulation has become better understood, more possible therapeutic strategies are emerging. For example, signaling through the IL-4 receptor is mediated by the activation of STAT6, so the development of STAT6 blocking agents is being actively pursued by several groups.

Several studies are ongoing to evaluate the action of anti-IgE antibodies in allergic rhinitis. Two different approaches have been proposed: producing antibodies to the membrane portion of IgE, but not secreted IgE, and developing antibodies against the IgE-binding site for the high-affinity IgE receptor Fc(RI). The first approach is predicated on the existence of a structurally different IgE isoform and is aimed at the ablation of surface IgE-bearing cells. The second relies on preventing IgE from binding to its receptor.

Finally, anti-CD4 monoclonal antibody has been reported to be beneficial in the animal models of allergic disorders.

SUMMARY

Allergic rhinitis is a very common disease, which is increasing in frequency throughout the developed world. Proper knowledge of disease pathogenesis and underlying immune mechanisms, as well as familiarity with current treatment guidelines, is essential for the proper differential diagnosis and management of allergic rhinitis. The latter should begin with simple environmental control measures to prevent the patient from inhaling potentially allergenic substances. After that, drug therapy should be instituted. Although intranasal corticosteroids are the most effective in this regard, oral and intranasal antihistamines and blockers of mast cell degranulation such as cromolyn can be used as first-line therapy, especially in children.

Initial referral to a specialist may be helpful for a majority of patients, to establish the diagnosis, identify and eliminate causative allergens, and develop a therapeutic plan. A limited response to pharmacotherapy and severe allergic symptoms most of the year are indications for allergen immunotherapy.

Suggested Readings

Beaven MA, Metzger H. Signal transduction by Fc receptors. Immunol Today 1993;14:222–226

Costa JJ, Galli SJ. Mast cells and basophils. In: Rich R, Fleicher TA, Shwartz BD, Shearer WT, Strober W, eds. Clinical Immunology: Principles and Practice. St. Louis: Mosby-Year Book; 1996: 408–430

Irani AA, Schechter NM, Craig SS, et al. Two types of human mast cells that have distinct neutral pretease compositions. Proc Natl Acad Sci U S A 1986;83:4464

Klinck M, Cline MG, Halonean M, et al. Problems in defining normal limits for serum IgE. J Allergy Clin Immunol 1990;85:440

Lemanske RFJ, Kaliner MA. Late phase allergic reactions. In: Middleton E Jr, Reed CE, Ellis EF, Adkinson NF, Yunginger JW, Buse WW, eds. Allergy: Principles and Practice, 4th ed. St. Louis: Mosby-Year Book; 1993:320–361

Malone DC, Lawson AA, Smith DH, Arrighi HM, Battista C. A cost of illness study of allergic rhinitis in the United States. J Allergy Clin Immunol 1997;99:22–27

Murray JJ, Tonel AB, Brash AR, et al. Release of prostaglandin D2 into human airways during acute antigen challenge. N Engl J Med 1986;315:800

Ogasawara H, Asakura K, Saito H, Kataura A. Role of CD4-positive cells in the pathogenesis of nasal allergy in the murine model. Int Arch Allergy Immunol 1999;118:37–43

Pipkorn U, Proud D, Lichenstein LM, et al. Inhibition of mediator release in allergic rhinitis by pretreatment with topical glucocorticosteroids. N Engl J Med 1987;316:1506

Sanderson CJ. Interleukin-5, eosinophils and disease. Blood 1992; 79:3101–3109

Valone FH, Boggs JM, Goetzl EJ. Lipid mediators of hypersensitivity and inflammation. In: Middleton E Jr, Reed CE, Ellis EF, Adkinson NF, Yunginger JW, Busse WW, eds. Allergy: Principles and Practice. 4th ed. St. Louis: Mosby-Year Book; 1993:302–319

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. The synthesis of immunoglobulin E (IgE), which plays a central role in allergic disease, is induced by the following:
   - A. IL-4, IL-5, IL-6, and IL-13, produced by Th2 cells
   - B. Interferon γ and TNF-β, produced by Th1 cells
   - C. IL-4, IL-12, and IL-13
   - D. IL-4 and IL-13, produced by Th2 cells
   - E. IL-4 and IL-13, produced by Th1 cells
2. Which of the following statements is true?
   - A. Histamine is the major mediator of an early-phase allergic reaction, and is contained in a preformed state within the granules of mast cells and basophils.
   - B. Prostaglandin D is formed in human mast cells from the metabolism of arachidonic acid through the cyclooxygenase pathway, and plays a role in maintaining acute and late allergic responses.
   - C. Leukotrienes are the products of arachidonic acid metabolism through the cyclooxygenase pathway, and they are the major mediators of late and chronic allergic reactions.
   - D. Tryptase, chymase, mast cell carboxypeptidase, and cathepsin G are located in both the mucosal mast cells and connective tissue mast cells.
   - E. The major proteoglycans in mast cells are histamine, heparin, and chondroitin sulfate A.
3. Which of the following statements regarding allergens is true?
   - A. Food and drug allergies are the common causes of allergic rhinitis.
   - B. The majority of plants in the northern United States produce windborne pollen and cause nasal allergy.
   - C. Almost all of the major mite allergens, which are the major cause of perennial allergic rhinitis throughout the world, are contained in mite fecal pellets.
   - D. Molds are highly ubiquitous outdoor allergens and usually cause only seasonal allergic rhinitis.
   - E. Ideal growing conditions for dust mites are low humidity, a temperature around 55°F, and the presence of shed human skin, which is their major food source.
4. Management of allergic rhinitis may include
   - I Environmental control
   - II Oral and intranasal antihistamines
   - III Intranasal steroids
   - IV Oral steroids
   - A. I, II, and III
   - B. I and III
   - C. II and IV
   - D. All the above

Chapter 4

HEAD AND NECK MANIFESTATIONS OF RHEUMATOLOGICAL DISEASES

Derek D. Sloan and Jeffrey P. Harris

This chapter will focus primarily on systemic rheumatological diseases that often significantly affect and can even initially present in the head and neck region. These include the collagen vascular or connective tissue diseases (e.g., rheumatoid arthritis and systemic lupus erythematosus) and the vasculitides (e.g., polyarteritis nodosa and Cogan's syndrome). There will also be a discussion of two organ-specific diseases (autoimmune thyroiditis and myasthenia gravis) and a brief overview of the head and neck manifestations of acquired immunodeficiency syndrome (AIDS). Preceding this review is a discussion of basic immunity, autoimmunity, and the unique immunology of the inner ear. The primary immunodeficiencies such as DiGeorge's syndrome (T-cell deficiency) and Bruton's agammaglobulinemia will not be covered in this chapter, nor will the association between upper airway malignancies and immunosuppression. Comprehensive reviews of these topics from an otolaryngologist's perspective have been covered by Harris and associates (Harris and Penn, 1981; Harris and South, 1982). Previous reviews have covered the head and neck manifestations of collagen vascular diseases (Standefer and Mattox, 1986) and autoimmune diseases (Campbell et al, 1983), as well as the association between systemic autoimmune diseases and hearing loss (Berrettini et al, 1998). Comprehensive reviews for head and neck surgeons of Wegener's granulomatosis (Devaney et al, 1998), scleroderma (Weisman and Calcaterra, 1978), giant cell arteritis and polymyalgia rheumatica (Ferguson et al, 1987), thyroiditis (Morinaka, 1995), and AIDS (Moazzez and Alvi, 1998; Riederer et al, 1996) are also available.

BASIC IMMUNITY

Although a comprehensive description of the underlying mechanisms that make up our immune system is beyond the scope of this chapter, an outline of the major components and their function is warranted.

There are two types of immunity: natural and acquired. Natural immunity, also called native or innate, refers to the defense mechanisms present prior to an individual's exposure to foreign molecules. This includes physicochemical barriers (mucous membranes), circulating molecules (complement), and phagocytic cells (macrophages and neutrophils). This discussion will emphasize the acquired immunity mediated by lymphocytes (T and B cells).

There are two types of acquired immune responses, humoral and cellular. Humoral responses are mediated by antibodies that are synthesized by B lymphocytes. Cellular responses are mediated by T lymphocytes. Antibodies are proteins found in the blood, lymphoid tissue, and mucosal tissue that specifically recognize and eliminate antigens. Antigens are molecules that elicit an antibody response. They include bacterial toxins such as lipopolysaccharides and viral proteins. Antigens are typically foreign molecules; however, this is not always the case (see Autoimmunity that follows). Although a humoral response can be transferred to a "naive" or unimmunized individual with cell-free portions of the blood (e.g., plasma or serum), an effective cellular response is required for a comprehensive humoral response. For example, an individual B cell clonally expands when the antibody on its surface recognizes its ligand (the specific binding site or epitope of an antigen). However, B cell proliferation and differentiation into a mature, secreting plasma cell require support from direct cell-to-cell interaction with accessory leukocytes (e.g., T helper cells and macrophages), as well as with soluble factors secreted by these cells.

The bone marrow is the site of origin of all lymphoid stem cell lines. The four primary classes of white blood cells (leukocytes) are lymphocytes, mononuclear phagocytes, dendritic cells, and granulocytes. Lymphocytes are divided into B cells, T cells, and natural killer cells. B cells continue to differentiate in the bone marrow and complete maturation in the peripheral circulation and lymphoid tissue. B cells are responsible for antibody production. T cells mature in the thymus and are divided into helper T cells, suppressor T cells, and cytotoxic T cells. Helper T cells stimulate B cells and are required for a comprehensive antibody response. They also activate macrophages and secrete cytokines. Cytotoxic T cells lyse virus-infected cells and tumor cells. Suppressor T cells modulate and downregulate the immune response. Natural killer cells perform antibody-dependent cellular cytotoxicity (ADCC). Mononuclear phagocytes include the bone marrow–derived circulating monocytes and their tissue-resident counterparts, macrophages. They phagocytose foreign material, present antigen to B and T cells, and secrete cytokines. Dendritic cells include the Langerhans cells and are important in the induction of an immune response. Finally, granulocytes are key players in the acute inflammatory response. Granulocytes include neutrophils or polymorphonuclear leukocytes (PMN), eosinophils, basophils, and the tissue-resident counterparts of basophils, histamine-containing mast cells. They respond to cytokines and antibody-coated material, leading to phagocytosis and the release of granules containing cytotoxic compounds (e.g., peroxides and superoxide anion). This oxidative burst can damage host tissue as well as foreign, invading cells.

There are three phases of immune responses: cognitive, activation, and effector. The cognitive phase describes the specific recognition of antigen by soluble antibodies, antibodies on the surface of B cells, or T-cell receptors. T-cell receptors are surface molecules on T cells that bind specific peptide sequences from cellularly processed antigens when expressed in association with self-surface molecules, major histocompatibility complex (MHC). Class I MHC is found on all cells and is recognized by cytotoxic T cells expressing clusters of differentiation (CD)8 surface molecules. Class II MHC is expressed only by antigen-presenting cells (APCs) and is recognized by helper T cells expressing CD4 on their surface. APCs include macrophages, B cells, dendritic cells, and endothelial cells.

The activation phase of an immune response is the proliferation and differentiation induced in lymphocytes following the recognition of antigen. This phase requires the presence of helper T cells and/or APCs. In the effector phase, activated lymphocytes function to eliminate antigen. For example, antibodies opsonize microbes, making them more attractive for phagocytosis. Antibodies also neutralize toxins and activate the complement cascade pathway leading to bacterial lysis.

Soluble protein hormones secreted by helper and accessory cells are crucial in the activation and effector phases. These molecules are collectively called cytokines or lymphokines [e.g., interleukin (IL), interferon, and tumor necrosis factor (TNF)]. Helper T cells initially secrete IL-2, which serves as a growth factor for activated B and T cells. In the presence of IL-12 from B cells, a subset of helper T cells develops called Th1. Th1 cells continue to produce IL-2 and also make interferon γ and TNF-β. A Th1 response is largely cellular, prompting the activation of macrophages, natural killer cells, and cytotoxic T cells. If helper T cells are initially exposed to IL-4 from mast cells, they differentiate into Th2 cells, making ILs-4, 5, 9, and 10. This milieu supports a humoral response. Other lymphokines increase vascular permeability and function as chemotactants, causing leukocyte migration. Macrophages secrete IL-1, which stimulates resting B and T cells, and TNF-α, which activates neutrophils and also functions as a broad-based inflammatory mediator. TNF-α produces fever and upregulates adhesion molecules on the endothelial cells of postcapillary high endothelial venules (HEVs). The adhesion molecules on endothelial cells are termed addressins (e.g., glycosylation-dependent cell adhesion molecule-1 (GlyCAM) and vascular cell adhesion molecule-1 [VCAM]). The complementary ligands on T cells are called homing receptors (e.g., selectins and integrins). This process of upregulation of adhesion molecules and homing receptors leads to the margination and recruitment of circulating leukocytes at the site of an immune response.

AUTOIMMUNITY

The hallmark of acquired immunity is the ability to distinguish between self and nonself. When this ability breaks down, autoimmunity results. A complex series of deoxyribonucleic acid (DNA) rearrangements and post-translational modifications in the variable regions of the immunoglobulin (Ig; e.g., antibody) and T-cell receptor genes results in the vast array of unique B and T cells we have at birth. It is estimated that we possess the ability to recognize over one billion different molecular targets via these unique variable regions. Essential to this process is the "education" or selection during fetal development in favor of cells that can recognize and thus be stimulated by their accessory counterpart cells (i.e., positive selection) and against cells that recognize self too strongly (i.e., negative selection). One theory that describes how an antigen can be a molecule from the host is through a loss of this negative selection. This is termed a loss of central tolerance.

Autoimmunity can also develop from a loss of peripheral tolerance called clonal anergy. Peripheral tolerance is the inhibitory effect that costimulatory-deficient APCs are thought to exhibit on peripheral T cells. With local infection or inflammation, APCs are stimulated and in turn can stimulate a resident T cell not specific for the original antigen. An example of this theory is the organ-specific disease autoimmune thyroiditis, which can be induced in animals administered thyroglobulin with a strong adjuvant.

As opposed to an organ-specific autoimmune disease, systemic autoimmune diseases such as lupus erythematosus and Sjögren's syndrome are thought to develop following polyclonal lymphocyte activation. Molecules such as bacterial lipopolysaccharides stimulate B cells irrespective of a specific antigen. Self-reactive B cells, which had been present but anergic, can be turned on in this situation. It also has been postulated that certain bacterial superantigens can activate polyclonal T cells by binding portions of the T-cell receptor and MHC on APCs, thus forming a stimulatory bridge in the absence of any processed peptide.

If a foreign antigen bears enough resemblance to a self-antigen (or autoantigen), a cross-reactivity can ensue. The classic example of this process is rheumatic fever and subsequent valvular heart disease. Following repeated infections with group A streptococcus, individuals develop antistreptococcal antibodies that also recognize myocardial proteins. This leads to a chronic inflammatory state that can scar the valves and cause stenosis.

Three possible mechanisms exist for tissue damage in autoimmune responses. First, autoantibodies can directly recognize host antigens, either natural or modified. The subsequent binding of the carboxyl terminus of the antibody (fragment crystallizable [Fc] portion) by Fc receptors on PMNs and macrophages leads to phagocytosis. Complement lysis is also initiated by exposed Fc. Second, antigen– antibody immune complexes from a distant site can deposit in filtering tissues (e.g., glomeruli) without binding any antigen, still causing tissue damage secondary to migration of PMNs and monocytes. Third, cytotoxic T cells can damage tissue after being recruited by the release of cytokines from activated APCs and helper T cells.

IMMUNOLOGY OF THE INNER EAR

The inner ear contains the immune cells and mediators necessary for an immune response, and the endolymphatic sac is apparently involved in immunoregulation because it contains a resident population of lymphocytes. The inner ear can also mount a primary immune response against exogenous antigen introduced into the perilymph. This response is as effective as peritoneal presentation and more effective than middle ear presentation in eliciting systemic immunity. Furthermore, when antigen was injected into the inner ear of an animal previously sensitized by systemic immunization, an inflammatory infiltration of immunocompetent cells, production of local antibody, and cochlear damage were observed. When the endolymphatic sac is surgically destroyed or the duct obstructed, there is reduced inflammation and cochlear damage, thus further emphasizing the sac's role. Interestingly, the lymphocytes that appear in the inner ear during an immune response do not originate from the endolymphatic sac, and the cochlea has no resident leukocytes. These cells have been shown to enter from the circulation via the spiral modiolar vein (SMV), which behaves as a high endothelial venule during a secondary immune response. The issue of autoimmune reactivity to inner ear tissue was first addressed in a guinea pig model, where immunization with heterologous (bovine) inner ear tissue led to hearing loss in roughly 30% of the animals, along with histopathological evidence of inner ear degenerative changes and a mononuclear cell infiltration (Harris, 1987). Sera from these animals contained immunoglobulins against a 68 kilodilation (kDa) inner ear protein, which has been shown to be present in 33% of progressive sensorineural hearing loss (PSNHL) patients and 30% of Meniere's patients but only 5% of controls (Gottschlich et al, 1995). The clinical entity autoimmune inner ear disease (AIED) has been described as an idiopathic bilateral PSNHL. Patients with AIED have a positive Western blot against 68 kDa protein and are often responsive to steroid treatment. An autoimmune etiology for AIED has also been suggested because AIED patients have an increased risk of developing a systemic autoimmune disease on follow-up.

RHEUMATOID ARTHRITIS

Rheumatoid arthritis (RA) is a chronic, systemic inflammatory disease with a typically insidious onset and progressive, fluctuating course of polyarthropathy leading to thickened synovial membranes and eventual damage of subchondral bone and cartilage. There is symmetrical involvement of small hand joints [proximal interphalangeal (PIP) joints and metacarpophalangeal (MCP)] joints, wrists, ankles, and elbows. In advanced forms of disease, there is subluxation and ulnar deviation of the MCP joints and swan neck deformity in the fingers. Pulmonary involvement includes diffuse interstitial fibrosis, parenchymal nodules, and pleural effusion. Splenomegaly is found in 10% of patients. Prevalence is roughly 1 to 3%, and onset is more common in women between 25 and 50 years of age. A pathognomonic feature, the rheumatoid nodule, is a necrobiotic granuloma usually found at subcutaneous sites subject to trauma (e.g., the extensor surface of the forearm). Laboratory findings include an elevated erythrocyte sedimentation rate (ESR) and positive rheumatoid factor (RF), which indicates the presence of antibodies reactive to the Fc portion of IgG molecules. Although a positive RF is a sensitive test, it is also found in other chronic inflammatory states such as lupus, tuberculosus, and sarcoid.

Otological manifestations of RA include an increased incidence of sensorineural hearing loss (SNHL), with reported prevalence rates varying from 29 to 48%. A prospective study of 45 RA patients found a mild SNHL in 44% of patients, 36% with bilateral involvement (Kastanioudakis et al, 1995). Audiometric analysis was predominantly normal. There was no correlation between hearing loss and age, sex, disease duration, systemic involvement, autoantibodies, or treatment regimen. It is important to note that no temporal bone studies have been reported confirming the relationship between inner ear dysfunction and RA. Conductive hearing loss rates varying from 13 to 38% have been reported in RA patients. Loosening of the transducer mechanism secondary to erosion of the ossicles by inflammatory synovitis is one proposed mechanism.

Laryngeal manifestations, especially cricoarytenoid involvement, are common in RA. They have been reported in 26 to 80% of patients investigated. In the acute phase synovitis and effusion can lead to dysphonia, hoarseness, odynophagia, and dysphagia. There can also be a sense of throat fullness (globus) and in severe cases dyspnea or stridor if limitation of vocal cord abduction via the posterior cricoarytenoid occurs. There have been numerous reports of an increased incidence of vocal cord nodules in RA patients, usually presenting as hoarseness and dysphonia. After cordotomy and removal of the subcordal masses, a patient's dysphonia typically resolves.

The temporomandibular joint (TMJ) is commonly involved in RA. TMJ arthritis can lead to symptoms of jaw, face, or ear tenderness and pain, malocclusion with limited jaw mobility (< 45 mm), and even dysphagia. TMJ involvement has been found in as many as 78% of RA patients evaluated with radiographic evidence that typically demonstrates flattening of the anterior portion of the condylar head. Additionally, RA can affect the nose, leading to nasal septum perforation. Arthritic involvement of the cervical spine in RA has a predilection for the atlantoaxial and apophyseal joints, and it has also been implicated in cervical myelopathy.

Treatment of RA is traditionally done with a medication pyramid. Nonsteroidal anti-inflammatory drugs (NSAIDs) are the first-line agent of choice. In intermediate RA or if early RA is refractory to NSAIDs, second-line agents such as methotrexate are often initiated. Other second-line agents include antimalarials, gold, D-penicillamine or sulfasalazine, and immunosuppressives such as cyclophosphamide and cyclosporine. Oral glucocorticosteroids can be substituted or added to the regimen usually at no more than 10 mg per day. Flares are often treated with pulse methylprednisolone boluses. Many studies have shown a benefit in combination therapy with disease-modifying antirheumatic drugs (DMARDs) such as methotrexate and cyclophosphamide or methotrexate and cyclosporine. New treatments include drugs that inhibit TNF-α receptors such as Etanercept, a soluble TNF receptor. This approach presumably works in part by preventing the activation of neutrophils and upregulation of adhesion molecules.

Highlights

Arthritis of the cricoarytenoid joints is very common in RA patients. Symptoms are usually hoarseness, dysphonia, and dysphagia; however, life-threatening upper-airway obstruction can occur if the cords become fixed in adduction. Vocal cord nodules are easily treated with cordotomy. TMJ arthritis is also a frequent finding, leading to jaw, face, and ear pain, as well as malocclusion and limited jaw mobility. Of note, one third to one half of RA patients have significant, demonstrable hearing loss. Sensorineural hearing loss is more common than conductive, and both unilateral and bilateral presentations are common. Novel drugs targeting the TNF-α receptor appear to be a promising treatment.

SYSTEMIC LUPUS ERYTHEMATOSUS

Systemic lupus erythematosus (SLE) is an inflammatory, multisystem disease affecting connective tissue, blood vessels, mucous membranes, and serosal surfaces. Onset may be sudden or insidious (often with fevers and malaise) and predominantly occurs in young women. Ninety percent of patients report articular symptoms (polyarthralgia, arthritis). The characteristic malar "butterfly" rash is pathognomonic but not common. Other findings include pleuritis, pericarditis, endocarditis, generalized adenopathy, splenomegaly, neuritis, meningitis, headaches, epilepsy, cerebral vascular accidents, scleritis, retinal degeneration, glomerulonephritis, and inflammatory bowel disease. Positive antinuclear antibodies (ANAs) occur in $> 98\%$ of patients with SLE; however, false-positives may be seen in 5 to 10% of cases. High titers of more specific anti-DNA antibodies (e.g., Farr test) can confirm an SLE diagnosis.

In a retrospective study of 673 patients with systemic vasculitides, 80 patients reported subjective audiological disturbances; of these patients, 14 had evidence of PSNHL by audiometry. Of the 69 patients diagnosed with SLE, there were no patients with PSNHL (Berrettini et al, 1998). However, others have reported SNHL in controlled studies of SLE patients as well as hearing loss with additional otologic manifestations such as chronic otitis media with necrotizing vasculitis and dysequilibrium. In a study of 84 patients with SLE, 26 patients (31%) reported aural symptoms. Fifty percent of these described unilateral hearing loss, and 50% bilateral with or without tinnitus. Although audiometry was offered to all subjects, only 10 completed testing, and seven of these had abnormal pure-tone thresholds (Sperling et al, 1998). Temporal bone histopathologic studies from SLE patients have shown evidence of new bone formation and fibrosis throughout the inner ear. A mechanism for these findings has been postulated using an animal model, where systemic immunization followed by local deposition of the same antigen in the inner ear led to inflammation of the spiral modiolar vein, leukocyte infiltration, cochlear fibrotic changes, and osteogenesis.

Head and neck manifestations have been reported in as many as 76% of patients with SLE by Rothfield (1981). This same author found that oral mucosal lesions, including palatal ulcers and hyperkeratosis, appeared as the first sign of disease in roughly 40% of patients. Laryngeal involvement in SLE has been reported with an incidence between 0.3 and 12.8%. Less frequent ear, nose, and throat (ENT) manifestations of SLE include TMJ arthritis, auricular chondritis in patients with associated relapsing polychondritis, and nasal septal perforations. Necrotizing tracheitis and parotid gland enlargement were found in roughly 10% of SLE patients in a different study. Additional findings (Ropes et al. [1976]) included a roughly 15% associated cranial neuropathy rate, with the trigeminal nerve and facial nerve most commonly involved.

Preventive treatment is critical with SLE and includes avoidance of sun exposure because of photosensitivity, routine laboratory screening to detect early renal and hematologic complications (e.g., chemistries, complete blood count, urinalysis, and immune studies), birth control, and infection monitoring. The initial pharmacotherapy, as with RA, is NSAIDs. Oral corticosteroids and antimalarials play an important role in long-term reduction of flares. Azathioprine, cyclophosphamide, and methotrexate are investigational drugs. Although cyclophoshamide has been shown to be more efficacious than azathioprine, azathioprine is considered safer. Side effects common to both drugs are gastrointestinal intolerance and bone marrow suppression. Numerous studies have demonstrated an advantage of combination oral prednisone and intravenous cyclophosphamide over oral prednisone alone in preventing nephritis. Intravenous corticosteroids are the drug of choice for acute flares.

Highlights

SLE will usually involve the head and neck region (three quarters of patients). Oral mucosal lesions are the most prevalent finding and are often the presenting symptom of disease. As many as one third of SLE patients will complain of aural symptoms, and SNHL can be demonstrated in the majority of those tested. Additional findings include trigeminal and facial nerve neuropathies, parotiditis, tracheitis, laryngeal manifestations, and TMJ arthritis.

PROGRESSIVE SYSTEMIC SCLEROSIS

Progressive systemic sclerosis (PSS) is a degenerative disease characterized by diffuse fibrosis and vascular changes in the skin (scleroderma), joints, and internal organs. There are two types: localized (most common in children) and systemic (usually manifest in adults). The incidence rate of systemic scleroderma is roughly 20 per million per year in the United States, and the prevalence is 240 per million. There is a 5:1 female to male ratio in PSS, and the 2-year survival rate is reportedly between 40 and 80%. Initial findings are commonly Raynaud's phenomenon (an intermittent pallor or cyanosis of acral parts, typically digits, secondary to vasospasm), polyarthralgia, gastrointestinal (GI) complaints (dysphagia, dyspepsia), and occasionally dyspnea. Progression can be rapid or prolonged before full manifestation of the CREST syndrome (*c*alcinosis of fingertips and bony eminences, *R*aynaud's, *e*sophageal dysfunction, *s*clerodactyly, and *t*elangiectasia). Visceral involvement includes GI (reflux with Barrett's metaplasia in one third of patients), cardiorespiratory (pulmonary fibrosis and cardiac failure), and renal systems (malignant hypertension and renal insufficiency).

In the Berrettini et al (1998) study of 673 patients with systemic vasculitis disease, 252 patients (37%) had systemic scleroderma. None of these patients had audiovestibular abnormalities as measured by audiometry, tympanometry, brainstem auditory evoked response (BAER), or electronystagmography (ENG); however, it is important to note that only those patients with subjective complaints were tested (80 of 673).

In a review of 71 PSS patients, head and neck manifestations were reported in 80% of those studied (Weisman and Calcaterra, 1978). Thirty percent of these patients had complaints in the head and neck region as the initial presentation. Dermatologic involvement included a tight, masklike facies secondary to a progression of the fibrosis (35%); telangiectasias of the face, lips, and tongue (18%); hyperpigmentation (7%); and calcinosis (3%). Investigators also found decreased opening of the mouth (28%), neck stiffness or pain (14%), xerostomia (8%), dysphonia (6%), keratoconjunctivitis sicca (2%), diminished taste (1%), and parotitis (1%). Additional oral mucosal changes in PSS include gingivitis, tongue atrophy, and thickening of the periodontal membrane.

Dysphagia, dyspepsia, and reflux are common complaints in PSS. Reported rates of dysphagia are roughly 40%, and abnormal esophagrams are present in 20% of PSS patients. An investigation of esophageal function in 125 scleroderma patients revealed that 45 patients (36%) had endoscopic evidence of esophagitis. Manometry revealed that 80% of those tested had abnormalities correlating with severity of disease (Bassotti et al, 1997).

Disease-modifying agents for PSS include drugs affecting vascular changes such as prostacyclin, drugs affecting immune response such as cyclosporin A, and drugs inhibiting fibroblast activity such as D-penicillamine and interferon-γ. Although D-penicillamine and interferon-γ are effective in slowing the rate of skin thickening, aggressive physical therapy, analgesics, and reconstructive surgery are also used. Esophageal dysmotility in PSS is managed like reflux disease from other causes with smaller meals, avoidance of late-night meals, eating upright, and histamine-2 receptor antagonist (H2) blockers or proton pump inhibitors. Dysphagia must be assessed with endoscopy to determine if esophageal strictures are present. Cisapride can be effective in improving esophageal muscle function. Cardiac and pulmonary complications often respond to NSAIDs or steroids, and angiotensin converting enzyme (ACE) inhibitors are a crucial tool used to control hypertension and prevent renal crisis in PSS patients.

Highlights

PSS involves the head and neck in 80% of patients and presents in this region 30% of the time. Esophageal dysfunction and the associated symptoms of dysphagia and reflux are perhaps the most common findings and can be easily demonstrated with manometry. Endoscopy should be done to reveal potential strictures and to assess for esophagitis. A tight, masklike facies, restriction of jaw mobility, telangiectasias, and neck pain and stiffness are also quite common.

POLYMYOSITIS/DERMATOMYOSITIS

Polymyositis is an inflammatory, degenerative myopathy producing proximal muscle weakness and tenderness principally in the shoulder and pelvic girdles. The manifestations of dermatomyositis include the pathognomonic periorbital edema with a heliotrope hue. Common findings also include polyarthralgia (one third of patients), interstitial pneumonitis producing dyspnea and cough, cardiac arrythmias, renal failure secondary to rhabdomyolysis, GI ulcerations producing hematemesis and melena (more common in children), and constitutional symptoms of fever, fatigue, and weight loss. There is an increased incidence of Raynaud's phenomenon and Sjögren's syndrome in patients with polymyositis. Onset is often acute in children and more gradual in adults, with the most common age of onset between 40 and 60 years of age. The prevalence is less common than SLE or PSS but more common than polyarteritis nodosa, and the female to male ratio is 2:1.

There are many findings in the head and neck region of polymyositis patients, including dysphonia, diffuse facial erythema, and stomatitis. Dysphagia and regurgitation from involvement of the striated muscle of the pharynx and upper esophagus are common occurrences, with reported rates as high as 84%, with manometric evidence of disease in 45% of these cases. Involvement of the neck flexors leading to profound weakness is found in approximately one fourth to one third of patients with idiopathic inflammatory myopathies, although facial weakness is much less common (2–5%).

Although evidence of audiovestibular manifestations in patients with polymyositis is not common in the literature (e.g., Berrettini et al, 1998, looked at 16 patients with polymyositis and found no evidence of PSNHL or vestibular pathology), there has been a reported association with autoimmune inner ear disease.

In addition to the increased association with other connective tissue diseases as previously mentioned, there is a roughly 15% associated malignancy rate. Neoplasms of the lung and breast are most common, but tonsillar and parotid carcinomas have also been reported, as has an increased incidence of nasopharyngeal carcinoma.

Corticosteroids are the mainstay of medical therapy in polymyositis. Prednisone is initiated at 40 to 80 mg per day for 1 to 2 months or until remission is achieved. Dosages are then tapered to the lowest possible levels to avoid the deleterious side effects of chronic steroids such as cataracts, osteoporosis, hypertension, skin fragility, poor wound healing and infection, weight gain, and hirsutism. Methotrexate has been used successfully in conjunction with steroids, as has azathioprine.

Highlights

The majority of polymyositis patients will complain of dysphagia and regurgitation with or without dysphonia. Neck flexor weakness is also common. There is an increased incidence of associated Sjögren's syndrome as well as carcinomas of the parotid gland and nasopharynx.

MIXED CONNECTIVE TISSUE DISEASE

Mixed connective tissue disease (MCTD), also called early undifferentiated connective tissue disease or overlap syndrome, has clinically similar characteristics to SLE, RA, PSS, and polymyositis. It is distinguished by the presence of high titers to nuclear ribonucleoprotein antigen. In MCTD there is also hypergammaglobulinemia, circulating immune complexes during disease flare, and deposition of antibody complexes within blood vessel walls and the glomerular basement membrane, although, unlike SLE, there is normal reticuloendothelial system (RES) clearance of immune complexes. Etiology is unknown, as is exact prevalence, but MCTD is probably more frequent than polymyositis and less frequent than SLE. The female to male ratio is 4:1, and age of onset has a mean of 37 years, with a range from 5 to 80 years reported. Raynaud's phenomenon is a common presentation, sometimes preceding other signs and symptoms by years. Polyarthralgia and arthritis are present in the vast majority of MCTD patients, with proximal muscle weakness being quite common. Skin manifestations include hand swelling with a sausage-like finger appearance. Eighty percent develop a pulmonary component, including interstitial disease and pleuritis. Pericarditis is the most frequent cardiac finding, although renal involvement is not as common (10%).

Highlights

Otologic involvement has not been reported in MCTD. The two most common head and neck manifestations are esophageal dysmotility with a decrease in upper sphincter pressure, present in 80% of patients, and an increased frequency of trigeminal sensory neuropathy. Persistent hoarseness and diffuse neck swelling secondary to pseudothrombophlebitis have also been described in MCTD patients.

SJÖGREN'S SYNDROME

Sjögren's syndrome is a chronic, inflammatory disease that affects the exocrine glands via lymphocytic infiltration. The resulting lacrimal and salivary gland atrophy in primary Sjögren's syndrome leads to the classic findings of dry eyes (keratoconjunctivitis sicca) and dry mouth (xerostomia). Secondary Sjögren's syndrome also includes systemic manifestations present in other collagen-vascular diseases (e.g., SLE, RA, and scleroderma). One third of Sjögren's patients have polyarthritis, with a similar distribution to RA, although with less joint destruction. Other findings are GI mucosal atrophy, pancreatic insufficiency, fibrinous pericarditis, renal tubular acidosis and interstitial nephritis, polymyositis, and cranial vasculitis. Sjögren's syndrome is more prevalent in females and more common than SLE but less common than RA. Diagnosis is confirmed by labial salivary gland biopsy and a Schirmer's filter paper test. Anti–Sjögren's syndrome A and B antibodies are prevalent (B antibodies being highly specific, although still found in SLE). Sjögren's patients have a 44-fold increased risk for lymphoma as well as an increased incidence of Waldenström's macroglobulinemia. These patients often have parotid enlargement, lymphadenopathy, and splenomegaly.

Otologic involvement in Sjögren's syndrome has been reported in the literature. In a cross-sectional study by Tumiati and Casoli (1997), 14 out of 30 Sjögren's patients (46%, 95%; confidence interval (CI) 28–66%) compared with one out of 40 controls (2.5%, CI 0.06–13%) had statistically significant ($p < .001$) evidence of SNHL. In another prospective study by Trott et al. (1996) investigating whether hearing loss in Sjögren's syndrome is immune related, three out of 14 Sjögren's patients had audiological evidence of SNHL. However, these three subjects had negative lymphocyte transformation tests (LTTs), whereas the two subjects with a positive LTT had no hearing loss. Further prospective studies with larger sample size and appropriate controls, as well as histological evidence, are needed to confirm an association between Sjögren's syndrome and immune-mediated hearing loss. It is possible that hearing loss in Sjögren's syndrome is an early manifestation of a concurrent, systemic inflammatory disease. For instance, a patient with a 6-year diagnosis of Sjögren's syndrome developed sudden SNHL and vertigo, with positive IgM and IgG antiphospholipid antibodies, suggesting an atypical Cogan's syndrome.

In addition to the otologic ramifications, there is substantial head and neck involvement in Sjögren's syndrome. Painless parotid gland enlargement is present in roughly one third of patients. Salivary gland atrophy and diminished saliva lead to xerostomia, as previously mentioned, but also increase the risk for dental caries, dysphagia, and dysphonia. Sjögren's syndrome patients with salivary gland dysfunction have significant swallowing impairment and increased pharyngeal transit time and complete fewer speech tasks relative to controls. Mucosal desiccation in the nose and throat increases the risk of infection, and although septal perforation is rare, nasal crusting and hyposmia are common findings. Finally, sensory neuropathy in Sjögren's patients occurs and is most common in the maxillary and mandibular branches of the trigeminal nerve.

Artificial tears provide symptomatic treatment for keratoconjunctivitis sicca. Xerostomia is more difficult to treat. It involves the palliative use of sugarless, flavorful lozenges, lemon juice, or gum, while avoiding smoking, dry foods, and drugs with anticholinergic effects. Stannous fluoride is used to prevent dental caries. Extraglandular disease is managed with systemic corticosteroids or immunosuppressive drugs such as cyclophosphamide.

Highlights

Obvious associations with Sjögren's syndrome and otolaryngology include the defining symptoms of keratoconjunctivitis sicca and xerostomia. Decreased salivary gland secretion leads to an increased risk of dysphagia, dysphonia, and dental caries. Painless parotid gland enlargement is seen in one third of patients. Perhaps more surprisingly, SNHL is nearly 20 times as common in Sjögren's patients.

BEHÇET'S SYNDROME

Beçhet's syndrome, also called oculo-oral-genital syndrome, is a relapsing, chronic inflammatory illness. It is characterized by recurrent oral aphthous ulcers (often the first manifestations and present in 100% of patients), plus two of these additional findings: painful genital ulcerations (80%), cutaneous vasculitic lesions (typically pseudofolliculitis or erythema nodosum–like reaction found in 60–80% of patients), synovitis and nondestructive arthritis (50%), uveitis or retinal vasculitis (60–70%), and meningoencephalitis (18%). Other characteristics include migratory thrombophlebitis occurring in 25% of patients, general vasculitis producing focal glomerulonephritis, and GI findings varying from abdominal pain to a regional enteritis resembling Crohn's disease. Presentation is usually in the third decade. The male to female ratio is 2:1, and disease prevalence is higher in some Mediterranean and Asian countries. There is an increased proportion of human leukocyte antigen (HLA)-B5 in Beçhet's patients, but there are no specific laboratory findings.

In the previously mentioned study by Berretini et al (1998), there were 36 patients with Beçhet's syndrome. Two of these patients complained of hearing loss, with audiological evidence of bilateral, asymmetrical SNHL. BAERs showed a cochlear pattern in both patients. One patient showed improved hearing after treatment with cyclophosphamide and methylprednisolone. This same patient complained of unsteadiness and on vestibular testing had evidence of defects in smooth pursuit and saccades as well as a right-directional preponderance without canal paresis. Although it was postulated that an AIED picture was the likely underlying etiology for hearing loss, there was no report of testing via the 68 kDa Western blot or LTT in these patients. Others have previously described audiovestibular disturbances in Beçhet's patients; however, the patient sample sizes were too small to demonstrate a definitive association.

Skin involvement in the head and neck region of Beçhet's patients is more common on the lips, tongue, buccal mucosa, soft and hard palates, tonsils, and pharynx but can also be seen on the nose. Ocular findings are present in most cases. They include relapsing iridocyclitis and retinal vasculitis and can lead to periorbital pain, hazy vision, photophobia, and occasionally blindness.

The risk of blindness in Beçhet's syndrome requires aggressive management. Azathioprine at 2.5 mg/kg per day has been shown to maintain visual acuity. Impressively, patients receiving azathioprine have better survival probability than placebo-treated patients even 6 years after a 2-year regimen of the drug. Cyclosporin A is an alternative to azathioprine, but nephrotoxicity, cost, and relapse with discontinuation limit its usefulness. Brief courses of systemic corticosteroids are used to treat acute flares involving the ocular system or other organs. Colchicine is also used to treat erythema nodosum and arthralgias, and sulfasalazine can be helpful with GI complications.

Highlights

All Beçhet's patients have recurrent oral aphthous ulcers. Cutaneous, vasculitic lesions can be found on the lips and face. There has been a reported association between bilateral SNHL and Beçhet's syndrome. The risk of blindness is the most serious complication. Immunosuppressive drugs such as azathioprine have been effective in managing this illness.

RELAPSING POLYCHONDRITIS

Relapsing polychondritis (RP) is an episodic, inflammatory disorder characterized by necrosis of cartilaginous structures and other connective tissues. Painful erythema and swelling of the auricles are the most common presentation. Typically, the inflammation of RP involves the pinna equally but spares the lobule, whereas bacterial perichondritis involves the entire pinna. The nose, larynx, and trachea, as well as the eyes, peripheral joints, skin, kidneys, heart valves, and blood vessels, can all be involved. In a study of 62 RP patients by Zeuner et al (1997), the median age of onset was 47 years. Ninety-four percent of patients had auricular chondritis, 57% had nasal involvement, 50% had ocular involvement, and 31% had respiratory system involvement. RP is rare, occurring in roughly equal frequency in men and women, and approximately one third of RP patients have associated diseases such as RA and SLE. Laboratory findings include an increased ESR and a biologically false-positive Venereal Disease Research Laboratory Slide Test (VDRL). The presence of autoantibodies to cartilage and type II and IX collagen suggests a humoral, autoimmune etiology.

Numerous reports document audiovestibular manifestations in RP, including vertigo and hearing loss present in 10 to 40% of patients. The type and presentation of hearing loss are highly variable and can be either bilateral or unilateral, conductive or sensorineural, of sudden or gradual onset, and slowly or rapidly progressive. There have even been case reports of RP that initially were presented as bilateral SNHL and tinnitus without the classic auricular inflammation.

The presence of autoantibodies, as previously mentioned, coupled with the demonstration of type II collagen in the tectorial membrane and otic capsule, provides a possible autoimmune mechanism for the inner ear pathology in RP. An alternate mechanism could be ischemia secondary to obliterative vasculitis of the labyrinthine artery and its branches. A temporal bone study from a patient with RP and sudden deafness demonstrated severe degeneration of the membranous labyrinth and fibro-ossification of the cochlear basal turn and lateral semicircular canal without evidence of endolymphatic hydrops. Conductive hearing loss in RP can occur either as a result of swelling and obstruction of the external ear canal or through direct involvement of the ossicular chain or eustachian tube.

Other findings, such as nasal chondritis, occur in as many as 75% of RP patients (Campbell et al, 1983). In some cases the destructive inflammation of the nasal dorsum leads to saddle nose deformity. The nasopharyngeal mucosa is typically spared, although mild dryness can be present. Ocular manifestations include conjunctivitis, episcleritis, uveitis, and iritis.

The degree of laryngotracheobronchial involvement varies. At the mild end of the spectrum, hoarseness, productive cough, sore throat, and dysphagia occur. The most life-threatening upper respiratory complication is airway collapse secondary to destruction of the supporting, cartilaginous tracheal rings.

Mild cases of RP can be treated with NSAIDs. Corticosteroids, such as prednisone at 30 to 60 mg daily, are initiated for patients with systemic or ocular symptoms. Immunosuppressives, such ascyclophosphamide, and azathioprine are used in conjunction with steroids. Cyclosporin A is sometimes used in refractory cases. Airway management is a critical issue in RP. Intravenous (IV) pulse methylprednisolone has successfully treated acute airway obstruction; however, emergent tracheostomy is sometimes necessary, and stents are used with tracheal collapse. Surgery is also indicated for patients with severe cardiac valvular involvement or large vessel aneurysms. Hearing loss associated with RP can respond to steroid therapy.

Highlights

Nearly all RP patients have the classical auricular chondritis sparing the lobule. Over one half to three quarters of patients have nasal chondritis, and one half of patients have ocular symptoms. Vertigo and variable forms of hearing loss are found in as many as 40% of patients and can be the presenting symptom. The most life-threatening complication of RP is airway compromise secondary to the destruction of tracheal rings.

WEGENER'S GRANULOMATOSIS

Wegener's granulomatosis (WG), first described by Friedrich Wegener in 1936, is characterized by localized, necrotizing, granulomatous inflammation in the upper and lower respiratory tract and kidney (focal glomerulonephritis). The third element in the classic triad is generalized vasculitis. WG is uncommon, with a reported incidence of a few cases per million per year, and there is no predilection for either sex. Average age of onset is between 20 and 40 years; presentation in childhood is exceedingly rare. Characteristic light microscopic findings include evidence of necrosis, granulomas, and vasculitis; however, neck biopsies often lack one or more of these features. Laboratory findings include an elevated ESR and high titers of antineutrophil cytoplasmic antibodies (ANCAs). Sensitivities for ANCAs in WG are between 90 and 95% when the disease is active and 60% when in remission. Although specificity for ANCAs is under investigation, coarsely granular cytoplasmic (c-ANCA) rather than perinuclear (p-ANCA) staining appears to be more specific for WG, and false-positives are found in polyarteritis nodosa, Kawasaki syndrome, and Sjögren's syndrome.

Devaney et al (1998) described a spectrum of head and neck manifestations in WG patients including nasal (pain, ulceration, edema, hyposmia, epistaxis, and deformity); sinus (headache and discharge); laryngeal (pain, dysphonia, stridor, wheezing, and dyspnea); otologic (otalgia, edema, and hearing loss); oral (ulceration and gingivitis); and ocular (pain, proptosis, and altered vision) involvement.

Middle ear findings, such as serous otitis media (SOM), are quite common in WG patients. Reported incidences are as high as 90%. Symptoms of conductive hearing loss, otalgia, otorrhea, and myringitis can result and have been reported as the presenting symptoms of disease (Fenton and Sullivan, 1994). Temporal bone analyses have shown evidence of granulation tissue around the eustachian tube and protympanum, providing evidence for the middle ear process that can lead to otitis media and conductive hearing loss.

SNHL has been found in as many as 81% of WG patients examined audiometrically. In an investigation by Berrettini et al (1998) of 15 patients with WG, two had bilateral, asymmetrical PSNHL, and one had bilateral mixed hearing loss. With cyclophoshamide treatment, one patient showed stabilization audiologically, one showed no change, and one worsened. Vestibular testing revealed that all three patients had bilateral canal paresis. One proposed mechanism for autoimmune SNHL involves a vasculitic destruction of the endolymphatic sac mucosa, perhaps similar to the immune-mediated process occurring in the mucosa of the respiratory and renal system in WG.

Not surprisingly, rhinitis and nasal obstruction are the most frequent head and neck presenting symptoms in WG patients. Additional nasal findings include hemorrhagic rhinorrhea and in severe cases saddle nose deformity; destructive lesions can even extend to the skull base. Sinusitis is a common finding, and computed tomographic evidence of paranasal sinus obliteration as well as orbital involvement has been reported. Additionally, WG can present as salivary gland enlargement, dysphonia, and obstructive airway symptoms secondary to laryngeal and subglottic destruction, and even facial nerve palsy.

Treatment of WG depends on the severity of organ involvement. More indolent forms of disease without renal complications (20% of patients) can often be managed with corticosteroids alone, and combined therapy with daily low-dose cyclophosphamide (2–4 mg/kg per day) is used in severe cases. In Guillevin in et al. (1997), a randomized, prospective trial comparing WG flares with steroids plus oral versus IV pulse cylcophosphamide found that the IV group had lower mortality rates and fewer complications over a 5-year study period. However, the oral cyclophosphamide regimen produced fewer relapses and longer remission periods. Combination therapy with oral prednisone and low-dose IV methotrexate led to a 60% remission rate in another study by de Groot et al. (1998), all immunosuppressive therapy regimens can lead to serious infectious complications such as *Pneumocystis carinii* pneumonia, and thus antibiotic prophylaxis is often warranted.

Highlights

Although nasal obstruction and rhinitis are probably the most common head and neck findings in WG, both SOM, resulting in conductive hearing loss, and SNHL occur frequently and have been reported as the initial presentation.

POLYMYALGIA RHEUMATICA AND GIANT CELL ARTERITIS

Polymyalgia rheumatica (PMR) and giant cell arteritis (GCA), also called temporal arteritis, are believed to be related entities, with roughly a 40% association between them. PMR usually presents with pain and stiffness of proximal muscle groups (neck, shoulder, and pelvic girdles). Stiffness is classically worse in the morning and after inactivity (gelling phenomenon). Polyarthralgias and synovitis also can occur, as can constitutional symptoms, including malaise, fever, depression, and weight loss. Onset is usually after 50 years of age, and the female to male ratio is 2:1. Incidence is roughly 54 per 100,000 population per year. A markedly elevated ESR is a sensitive but nonspecific marker, and muscle biopsies and electromyography (EMG) are normal. The most common ENT manifestation in PMR is neck pain and stiffness.

GCA is often classified with WG as a granulomatous vasculitis. Biopsies reveal the presence of granulomatous inflammation in the intima and inner media of large, elastic arteries (particularly the carotid and cranial system) with predominant lymphocytes and giant cells. Symptoms arise in the distribution of affected vessels secondary to arteritis and ischemia from lumen narrowing. Characteristics include temporal and occipital headaches, visual changes (amaurosis fugax, diplopia, scotoma, ptosis, and blindness), and scalp tenderness. Jaw claudication is also common, involving temporalis, masseter, and glossus muscles. Onset is after 50 years of age, with a prevalence of 1 per 1000 and an incidence of ~1 in 10,000 per year. Like PMR, most patients have an elevated ESR, and a normocytic-normochromic anemia is often present. A characteristic temporal artery biopsy will often confirm the diagnosis. In a review of GCA by Ferguson et al (1987), myalgias, headaches, and temporal tenderness were reported to occur in >50% of patients. Visual findings, anorexia, jaw claudication, scalp tenderness, fever, and facial pain occurred in 10 to 50%, and blindness, tongue claudication, stroke, angina, or myocardial infarction in <10%. Blindness will occur in up to one third of untreated GCA patients, and thus high dose steroids are the gold standard in preventive treatment. Other findings in the distribution of the temporal artery include alopecia, redness, and even necrosis of the scalp. Swelling of the temporalis muscles without pain also can occur. As previously mentioned, cranial arteritis can lead to jaw and tongue claudication, as well as trismus. It also can produce dysphagia, facial neuropathy, cough, and dysphonia.

There have been several case reports of acute onset, bilateral SNHL in GCA patients. The hearing loss can even be the presenting symptom of this disease, and it is usually steroid responsive. In the report of 17 patients with GCA by Berrettini et al (1998), only one complained of hearing impairment and unsteadiness. She had evidence of bilateral serous effusions with a type B tympanometric shape, and the audiogram showed a bilateral, mixed hearing loss. Vestibular exam showed bilateral canal palsy. Once the diagnosis of GCA was made by temporal artery biopsy, the patient was started on 6-methylprednisolone. The air–bone gap disappeared, but the SNHL persisted.

Recommended initial daily dosages of corticosteroids for treating PMR are typically 10 to 20 mg of prednisolone. A higher dosage of 40 to 60 mg of prednisolone is used for GCA, given the risk of blindness.

As previously mentioned, associated hearing loss, like blindness, typically responds to steroids.

Highlights

Neck pain and stiffness are the most common head and neck manifestations of PMR. In GCA patients, temporal tenderness and headaches occur most frequently, followed by visual changes, jaw claudication, facial pain, and tongue claudication. Blindness is the most serious complication. It will occur in roughly one third of patients not treated with corticosteroids. Bilateral SNHL can occur as the presenting symptom and responds to steroids.

POLYARTERITIS NODOSA

Polyarteritis nodosa (PAN), also known as periarteritis nodosa, is a systemic, vasculitic illness characterized by inflammation and necrosis of small to medium-sized muscular arteries, resulting in ischemic tissue damage. The renal system is most commonly affected. A rapidly progressive glomerular nephritis can lead to hypertension, nephrotic syndrome, and renal failure. Cardiovascular involvement, present in 70% of PAN patients, includes coronary arteritis leading to angina and myocardial infarction. Pericarditis can also occur. Gastrointestinal findings include abdominal pain, intestinal bleeding, and obstruction. Hepatic involvement ranges from mild tenderness and capsular vasculitis to massive, hepatic infarction. The neurological sequela is classically a mononeuritis multiplex affecting motor and sensory nerves in a bilateral, asymmetric fashion. Central nervous system (CNS) findings are typically headache, convulsions, and encephalopathy. Skin findings are uncommon but include the characteristic subcutaneous nodule. Laboratory findings are leukocytosis, microscopic hematuria, and proteinuria. The peak incidence of onset for PAN is 40 to 60 years old; the female to male ratio is 2.5:1. Presenting symptoms in decreasing order of frequency are fever, abdominal pain, mononeuritis multiplex, weight loss, and weakness.

Hearing loss has been reported in PAN and can be the initial presenting complaint of this systemic disease (Wolf et al, 1987). However, a purely perceptive hearing loss is not common. Temporal bone studies of PAN patients demonstrate a mechanism for SNHL by showing evidence of ischemic inner ear involvement, including arteritis of the internal auditory artery and osteogenesis and fibrosis of the basilar turns. Additional temporal bone studies have demonstrated a complete absence of the organ of Corti and stria vascularis, collapse of Reissner's membrane, and distortion of the tectorial membrane, as well as osteogenesis and fibrosis of the apical turns. Given the ability to reproduce these changes in animal models by introducing antigen into the inner ear or by microembolism of the internal auditory artery, it is reasonable to search for an underlying vasculitis as a recognizable cause of idiopathic, profound hearing loss.

Often the hearing loss in PAN is secondary to middle ear effusion and can appear as a purely conductive or mixed deficit (Wolf et al, 1987). Chronic middle ear changes have been seen in temporal bone studies of PAN patients demonstrating SOM with inflammatory cell infiltrates and thickened, plugged vessels surrounding the facial nerve. In Yoon et al. (1989) and Steenerson (1986), a facial nerve palsy has also been reported as the initial symptom of PAN. Additional cranial neuropathies and peripheral nerve involvement affecting the head and neck, such as Horner's syndrome, have been reported in PAN, although less frequently. Given the systemic nature of this illness, it is surprising how infrequently lesions are found in the oral and nasal mucosa.

Two conditions with similarity to PAN are Churg-Strauss syndrome and Kawasaki syndrome. Churg-Strauss is characterized by the classic triad of systemic vasculitis, asthma, and peripheral eosinophilia. In one report, 70% of patients with Churg-Strauss had nasal involvement, including obstruction, rhinorrhea, polyposis, crusting, and, less frequently, anterior septal perforation. Kawasaki syndrome (mucocutaneous lymph node syndrome) is a rare entity that primarily affects infants and young children but can occur in young adulthood. Fever, conjunctivitis, mouth and lip changes, atypical rash with desquamation, and cervical adenopathy are all common findings. A cardinal sign is red, xerotic lips and erythema of the oral mucosa, found in 90% of patients within the first week of disease. A prominence of the tongue papillae (strawberry-like appearance) and mucosal necrosis also can occur.

Limited, nonprogressive PAN is managed with oral daily prednisone. Without treatment, the 5-year mortality rate is 85%. With treatment, this decreases to 30 to 45%. Refractory disease is treated with combined steroid and cytotoxic therapy, including cyclophosphamide, azathioprine, and methotrexate, which appear to decrease the relapse rate. Plasmapheresis is used but has not been shown to be effective. Anti-CD4 monoclonal antibody therapy has also helped some PAN patients achieve remission, although this treatment is still in its infancy. IV immunologlobulin has been used in Kawasaki syndrome to decrease the rate of coronary aneurysms.

Highlights

The vasculitis in PAN patients most commonly affects the kidneys (hypertension), heart (angina), gastrointestinal system (abdominal pain), and nervous system (mononeuritis multiplex). Head and neck involvement is surprisingly uncommon, but symptoms such as a conductive or mixed hearing loss and facial nerve palsy do occur and can be the initial presentation.

COGAN'S SYNDROME

Cogan's syndrome is a disease of young adults characterized by vestibuloauditory abnormalities and a nonleutic, interstitial keratitis (IK). Patients develop acute attacks of vertigo, tinnitus, and SNHL usually 1 to 6 months before or after IK findings. If 2 years or more separate the audiovestibular findings from manifestations of IK, or if signs of episcleritis, uveitis, or conjunctivitis occur, a diagnosis of atypical Cogan's syndrome is considered. The IK can be associated with upper respiratory infection (URI) symptoms and causes sudden onset of pain, lacrimation, and photophobia, which gradually resolve with episodic flares. However, the inner ear involvement, often clinically indistinguishable from Meniere's disease, can progress to complete deafness in 1 to 3 months or leave the patient with varying degrees of fluctuating audiovestibular symptoms if left untreated. Cogan's syndrome is often associated (>50% of patients) with a generalized vasculitis, frequently PAN, leading to systemic symptoms of arthritis, splenomegaly, glomerulonephritis, hypertension, and inflammatory bowel disease. Angitis of the cerebral nerves as well as fatal aortic valvular disease can occur.

Numerous temporal bone studies have been done on Cogan's patients. Common findings are vasculitis, endolymphatic hydrops, and fibro-ossification of the perilymphatic spaces. Additional findings include plasma cell and lymphocytic infiltration of the spiral ligament, saccular rupture, osteogenesis of the round window, degeneration of the spiral ganglion and stria vascularis, destruction of the organ of Corti, and ectopic bone within the semicircular canals.

One postulated etiology of Cogan's syndrome involves a hypersensitivity response to an infectious agent associated with vasculitis. Lymphocyte transformation on exposure to scleroprotein and inner ear antigen also has been described. Using indirect immunofluorescence, investigators have demonstrated IgG antibodies to human inner ear vessels (stria vascularis and lamina spiralis ossea) and IgA and IgG antibodies to human cornea in the serum of Cogan's patients. These reports give credence to an autoimmune mechanism, but, as we have previously discussed, it is unknown whether autoantibodies and activated lymphocytes are the primary event leading to tissue damage or merely a response to immune-stimulating antigen released from a non-organ-specific event (e.g., immune complex–mediated vasculitis).

Harada syndrome, also known as Vogt-Koyanagi-Harada (VKH) syndrome, has similarities to Cogan's syndrome. SNHL, dizziness and vertigo, granulomatous uveitis, depigmentation of the hair and skin around the eyes, and loss of eyelashes are all common features. A central type of VKH syndrome involving the CNS, typically aseptic meningitis, is less frequent than the peripheral type and when present helps distinguish this illness from Cogan's syndrome. The etiology of VKH syndrome is thought to involve an autoimmunity to melanocytes in the inner ear, uvea, skin, and meninges.

IK from Cogan's syndrome readily responds to topical and systemic corticosteroids. The prognosis for reversal of hearing loss and vestibular symptoms is greatly improved if steroid therapy is begun within 2 weeks of the onset of symptoms.

Highlights

Cogan's patients present with acute attacks of fluctuating audiovestibular symptoms (SNHL, tinnitus, vertigo), which can mimic Meniere's disease. Corneal eye involvement (IK) follows within 6 months in Cogan's syndrome, and one half of patients develop a systemic vasculitis. Aural and ocular symptoms respond to steroids but must be initiated soon after presentation.

SERONEGATIVE SPONDYLOARTHROPATHIES

The seronegative spondyloarthropathies, previously known as rheumatoid variants, all have a genetic disposition for the HLA-B27 antigen. This group includes ankylosing spondylitis, psoriatic arthritis, Reiter's disease, and the enteropathic arthritides. Inflammation of the spine commonly begins in the lumbosacral region and extends to involve the cervical spine; however, some patients may have isolated cervical involvement. Syndesmophytes, new bone proliferation in response to the spondylitis, bridge vertebral bodies and can lead to fusion, inflexibility, and an increased risk of fracture (Campbell et al, 1983).

Reiter's disease has significant manifestations in the head and neck. It has been described by the triad of arthritis, conjunctivitis, and urethritis. This disease usually affects young males; however, it is more frequently being recognized in females. Oral involvement is quite common in Reiter's disease. It has been reported in roughly 40% of patients studied. Lesions typically are painless and superficial and can occur anywhere in the oral mucosa. Tongue lesions can begin as vesicular eruptions with an erythematous base that then progress to ulcerations. Diffuse stomatitis also may be present. Oral lesions usually regress spontaneously within a few days of presentation.

AUTOIMMUNE THYROIDITIS

Autoimmune thyroiditis can manifest as hyperthyroidism (Graves' disease) or as hypothyroidism (Hashimoto's thyroiditis). In Graves' disease autoantibodies are directed against the thyroid-stimulating hormone (TSH) receptor thyroid-stimulating immunoglobulins (TSI). Symptoms and signs of different causes of hyperthyroidism are indistinguishable. They include warm, moist skin, nervousness, weight loss, tachycardia, hypertension, heat insensitivity, and frequent bowel movements. Graves' disease is characterized by the classic triad of diffuse toxic goiter, infiltrative ophthalmopathy, and infiltrative dermopathy (pretibial myxedema). The ophthalmopathy can range from periorbital puffiness to extraocular muscle dysfunction, proptosis, exophthalmos, optic neuritis, and even blindness. Graves' disease affects women 5 to 10 times more frequently than men and peaks in the third and fourth decades. The course typically undergoes spontaneous remissions and relapses. The diagnosis is made by elevated thyroxine (T4), decreased TSH, and the presence of TSI. Well-differentiated thyroid cancer is 2.5 times more likely in Graves' patients compared with the general population. There are multiple therapeutic options in treating hyperthyroidism. Iodine suppresses the release of triiodothyronine (T3) and T4. Propylthiouracil and methimazole inhibit the organification and coupling reaction. Beta-adrenergic blockers decrease the symptoms and signs of disease. Ablative therapy with radioactive iodine-131 is a permanent solution for women past their childbearing years, and surgery is an alternative for patients under 21 years old, those with goiters greater than 100 g, and for some patients with toxic adenoma or multinodular goiter.

Hashimoto's thyroiditis is the most common type of thyroiditis and is six times more common in women, peaking in the fifth and sixth decades. It is seen in association with systemic autoimmune diseases such as RA, SLE, and Sjögren's syndrome. Common hypothyroidism symptoms include dry, thick skin, fatigue, edema, cold intolerance, weight gain, menstrual disorders, depression, and constipation. Head and neck symptoms include enlarged tongue, hoarseness, blurred vision, and middle ear effusion. Typically, the thyroid gland is large, firm, and lobulated but can be nonpalpable secondary to fibrotic changes. Hearing loss in Hashimoto's thyroiditis can be sensorineural, conductive, or mixed and occurs more frequently and more severely in children. A progressive mixed hearing loss was reported in over half of the children with endemic cretinism and in less than 40% of the adults studied. Vertigo and tinnitus are also common, occurring in as many as 25% of patients reviewed. The treatment for Hashimoto's thyroiditis is lifelong replacement of endogenous thyroid hormone, usually accomplished with 100 to 150 μg of levo-thyroxine per day.

In a study by Morinaka (1995) of 6348 outpatients in an otolaryngology department over a 5-year period, investigators found that 87% of the 114 patients with thyroid disease had no subjective symptoms, 54% of the patients diagnosed with thyroid disease had Hashimoto's thyroiditis, and 2% had Graves' disease. Only five patients had visible goiter; however, 93% had palpable thyroid glands, thus emphasizing the importance of the physical neck exam.

MYASTHENIA GRAVIS

Myasthenia gravis is an autoimmune disorder affecting the neuromuscular junction and causing increasing muscle weakness with effort. Its prevalence is estimated at 2 to 10 per 100,000 patients. Twenty percent of patients experience onset of symptoms by age 20, and the majority of patients are affected by age 40. The pathology results from the action of antibodies directed against acetylcholine receptors (AChR Ab). The most common presenting symptoms are ptosis, diplopia, and muscle fatigability after exercise. Any striated muscles can be involved; however, 85% of patients will develop ocular involvement at some point. Oropharyngeal muscle involvement is quite common, producing dysarthria and dysphagia. It has been reported as the presenting symptom in 6 to 17% of myasthenia gravis patients studied (Sanders and Howard, 1991).

Anticholinesterases, such as pyridostigmine and neostigmine, and plasmapheresis are used to treat current symptoms in myasthenia gravis patients. Corticosteroids and azathioprine are used in semichronic management.

ACQUIRED IMMUNODEFICIENCY SYNDROME (AIDS)

Most individuals with AIDS will develop head and neck manifestations. In a prospective study, Riederer et al (1996) reported otologic findings in 250 AIDS patients, including otitis externa (6%) and otitis media with effusion (9%). Kaposi's sarcoma, eustachian tube dysfunction (ETD), and SNHL have also been reported, and estimates for SNHL in HIV patients range from 20 to 50%. Nasal-sinus and oral manifestations are even more common than otologic findings. Symptoms include cutaneous Kaposi's sarcoma, herpes simplex virus and zoster ulcers, and sinusitis (24%). Oral findings range from candidiasis, oral hairy leukoplakia, gingivitis, stomatitis, and peridontitis to non-Hodgkin's lymphoma and squamous cell carcinoma. Neck findings include generalized lymphadenopathy, parotid gland cysts, and neck masses due to infections with *Mycobacterium tuberculosis* and *Mycobacterium avium*, as well as fungal infections such as cryptococcosis, histoplasmosis, and coccidioidomycosis.

The treatment of HIV and AIDS is beyond the scope of this chapter and changes frequently. Proper management by a physician who has developed a good relationship with the patient and can provide counsel on the most current antiretroviral combinatory therapy as well as antibiotic prophylaxis is critical.

SUMMARY

The important role of the otolaryngologist in diagnosing and managing the diseases described above becomes apparent when reviewing the frequency of head and neck manifestations reported. Indeed, these findings are quite often the initial presentation of uncommon illnesses with nonspecific systemic symptoms. This leads to delayed diagnosis by the primary care physician and further underscores the essential contribution a head and neck surgeon can make when a patient is referred to his or her clinic. For example, RA, a relatively common disease not classically examined from an otolaryngologist's perspective, not surprisingly has a high association of laryngeal and TMJ involvement; however, as many as half of RA patients also have significant hearing loss. When working up a patient with sudden hearing loss or acute hoarseness and dysphagia, consider an underlying autoimmune etiology. If no other cause can be identified, consider obtaining laboratories such as an ESR, RA, and ANA. Prompt recognition and treatment of diseases such as Cogan's syndrome and WG can even prevent or reverse hearing loss. In cases where the head and neck symptom(s) begin(s) to dominate the illness, the otolaryngologist must also assume an integral component of the responsibility in managing the patient's care and work with other specialists such as rheumatologists and immunologists. Having a basic working knowledge of the fundamental pathophysiological processes of these diseases can help in the clinician's understanding and aid in the ability to judge future diagnostic and therapeutic changes.

Suggested Readings

Bassotti G, Battaglia E, Emanuelli G, et al. Esophageal dysfunction in scleroderma: relationship with disease subsets. Arthritis Rheum 1997;40(12):2252–2259

Berrettini S, Ferri C, Ravecca F, et al. Progressive sensorineural hearing impairment in systemic vasculitides. Sem Arthritis Rheum 1998;27(5):301–318

Campbell SM, Montanaro A, Bardana EJ. Head and neck manifestations of autoimmune disease. Am J Otolaryngol 1983; 4:187–216

Devaney KO, Ferlito A, Devaney S, et al. Wegener's granulomatosis of the head and neck. Ann Otol Rhinol Laryngol 1998; 107:439–445

Fenton JE, Sullivan TJ. The otological manifestations of Wegener's granulomatosis. J Laryngol Otol 1994;108:144–146

Ferguson BJ, Allen NB, Farmer JC. Giant cell arteritis and polymyalgia rheumatica: review for the otolaryngologist. Ann Otol Rhinol Laryngol 1987;96:373–379

Gottschlich S, Billings P, Harris JP, et al. Assessment of serum antibodies in patients with rapidly progressive sensorineural hearing loss and Meniere's disease. Laryngoscope 1995;105: 1347–1352

de Groot K, Muhier M, Reinhold-Keller E, et al. Induction of remission in Wegener's granulomatosis with low dose methotrexate. J Rheumatol 1998;25(3):492–495

Guillevin L, Cordier JF, Lhote F, et al. A prospective, multicenter, randomized trial comparing steroids and pulse cyclophosphamide versus steroids and oral cyclophosphamide in the treatment of generalized Wegener's granulomatosis. Arthritis Rheum 1997;40(12):2187–2198

Harris JP. Experimental autoimmune sensorineural hearing loss. Laryngoscope 1987;97:63–76

Harris JP, Penn I. Immunosuppression and the development of malignancies of the upper airway and related structures. Laryngoscope 1981;91:520–528

Harris JP, South MA. Immunodeficiency diseases: head and neck manifestations. Head Neck Surg 1982;5:114–124

Kastanioudakis I, Skevas A, Danielidis V, et al. Inner ear involvement in rheumatoid arthritis. J Laryngol Otol 1995;109(8):713–718

Moazzez AH, Alvi A. Head and neck manifestations of AIDS in adults. Am Fam Physician 1998;57(8):1813–1822

Morinaka S. On the frequency of thyroid disease in outpatients in an ENT clinic. Auris Nasus Larynx 1995;22(3):186–191

Riederer AP, Grein GO, Bogner JR. High prevalence of opportunistic infections in the head and neck related to human immunodeficiency virus: a prospective study of the distribution of otorhinolaryngologic disorders in 250 patients. Infection 1996;6:440–446

Ropes MW. Systemic lupus erythematosus. Cambridge: Harvard University; 1976:28–29

Rothfield N. Clinical features of systemic lupus erythematosus. In Kelly WN et al. (eds) Textbook of Rheumatology. Philadelphia: WB Saunders; 1981;1106–1132

Sanders DB, Howard JF. Disorders of neuromuscular transmission. In: Bradley W et al, eds. Neurology in Clinical Practice. Vol 2. London: Butterworth-Heinemann; 1991: 1819–1842

Sperling NM, Tehrani K, Ginzler E et al. Aural symptoms and hearing loss in patients with lupus. Otolaryngol Head Neck Surg 1998;118(6):762–765

Standefer JA, Mattox DE. Head and neck manifestations of collagen vascular diseases. Otolaryngol Clin North Am 1986; 19(1):181–210

Steenerson RL. Bilateral facial paralysis. Am J Otol 1986;7(2): 99–103

Trott MS, Hughes GB, Calabrese LH, et al. Hearing and Sjögren's syndrome. Ear Nose Throat J 1996;75(10):666–668

Tumiati B, Casoli P, ParmeggianiA. Hearing loss in Sjögren's syndrome. Ann Intern Med 1997;126(6):450–453

Weisman RA, Calcaterra TC. Head and neck manifestations of scleroderma. Ann Otol 1978;87:332–338

Wolf M, Kronenburg J, Engelberg S, Leventon G. Rapidly progressive hearing loss as a symptom of polyarteritis nodosa. Am J Otolaryngol 1987;8:105–108

Yoon TH, Paparella MM, Schachern PA. Systemic vasculitis: a temporal bone histopathologic study. Laryngoscope 1989; 99(6 Pt 1):600–609

Zeuner M, Straub RH, Rauh G, et al. Relapsing polychondritis: clinical and immunogenetic analysis of 62 patients. J Rheumatol 1997;24(1):96–101

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. Autoimmunity is thought to develop for the following reasons:
   A. Loss of peripheral tolerance and clonal anergy
   B. Polyclonal lymphocyte activation
   C. Loss of negative selection
   D. All of the above
2. The inner ear has become damaged in the following diseases except
   A. Autoimmune inner ear disease
   B. Relapsing polychondritis
   C. Mixed connective tissue disease
   D. Polyarteritis nodosa
3. The following are true regarding Western blot assay in autoimmune inner ear disease (AIED) except
   A. It is an assay involving the patient's sera reacting against a bovine 68 kDa antigen from the inner ear.
   B. Meniere's disease has been found to be positive in ~30% of cases.
   C. It has never been shown to be present in experimental autoimmune inner ear disease in animals.
   D. If positive, it is useful in predicting steroid responsiveness.
4. Head and neck manifestations of autoimmune disease include all of the following true statements except
   A. Immunosuppressive drugs are the mainstay of treatment for these illnesses.
   B. The inner ear fluid compartments when involved by autoimmune disorders develop fibrosis and osteoneogenesis.
   C. Though bothersome, these diseases are rarely life threatening and can be followed until their symptoms warrant the use of these dangerous drugs.
   D. Symptoms in these patients often wax and wane with activity of the systemic disease.

Chapter 5
PULMONARY PHYSIOLOGY AND MECHANICAL VENTILATION

Karen B. Zur and Gregory J. Schilero

Oxygen is the principal substrate for, and carbon dioxide the major by-product of, aerobic metabolism. To meet the energy needs of the body, tissue oxygen stores must constantly be replenished, coupled with the removal of carbon dioxide that is necessary for acid-base homeostasis. The lungs' principal function is to facilitate this exchange of gas, beginning with inspiration of oxygen-rich ambient air through the upper airway (mouth, pharynx, and larynx) into the tracheobronchial tree. Passage of air by bulk flow from the trachea to the level of the terminal bronchioles, termed the conducting zone, occurs through a succession of ~16 airway branch points or generations. Thereafter, gas is carried primarily by diffusion across another seven generations of airways comprising the pulmonary acini or lobules. The pulmonary lobules are anatomically defined by a primary respiratory bronchiole (the transition zone for gas exchange) in communication with alveolar ducts and sacs (the respiratory zone). At this level, an estimated 300 million alveoli bathed in pulmonary capillary blood provide an enormous surface area (estimated at 50 to 100 m^2) for oxygen uptake by the circulation and the removal of carbon dioxide to the atmosphere. In addition, the alveolar-capillary membrane under normal conditions is extremely thin (0.5μ) offering little resistance to the diffusion of gas.

Pulmonary physiology therefore deals primarily with the mechanics of breathing, the tidal exchange of air between the lungs and atmosphere (ventilation), pulmonary capillary blood flow (perfusion), ventilation–perfusion interchange, and gaseous diffusion. The complex interplay of these processes results in normal gas exchange without excess cardiac or pulmonary work. Even in the presence of extensive lung disease, the "pulmonary reserve" may be large enough and the mechanisms that adjust regional ventilation and blood flow within the lung so efficient that gas tensions remain within normal limits.

This basic review of pulmonary physiology will stress the mechanics of breathing, including functional aspects of the extrathoracic airway germane to the otolaryngologist. The description and measurement of lung volumes, pulmonary function testing, and patterns of impairment, and the fundamentals of gas exchange will also be discussed. The final section provides an overview of commonly used modes of mechanical ventilation.

LUNG VOLUMES: MEASUREMENT AND SIGNIFICANCE

There are four standard lung volumes and four capacities, the latter consisting of the summation of two or more standard volumes. Many of the subdivisions of lung volume can be determined using a spirometer, a simple device that measures the forced expiratory volume beginning from maximal inspiration plotted graphically against the expiratory time in seconds (**Fig. 5–1**). The total volume of air that can be expired following such a maneuver is termed the forced vital capacity (FVC), which is the sum of the expiratory reserve volume (ERV), tidal volume (TV), and inspiratory reserve volume (IRV). The inspiratory capacity (IC), in turn, is the sum of the TV and IRV. Another routine spirometric measurement is the forced expiratory volume in 1 second (FEV-1), which is the volume of air expired during the first second following a forced expiratory maneuver from maximal inspiration. Expiratory reserve volume is the maximal volume of gas that can be expelled from the resting end-expiratory position. Any condition that leads to increased intra-abdominal pressure (e.g., obesity, pregnancy) can result in a decrease in ERV.

Tidal volume, also measured by spirometry, is the volume of gas exchanged during quiet breathing. Viewed alone, TV is of minimal significance because it is variable even in normal subjects. In mechanically ventilated subjects on volume-cycled ventilators, however, the level of the preset TV assumes importance depending on the underlying pulmonary disease. As a weaning criterion for mechanically ventilated patients, a TV of 5 cc/kg is sometimes employed as a predictor of weaning success.

The functional residual capacity (FRC) is the volume of gas remaining in the lungs at the end of tidal expiration. Often referred to as the resting point of the respiratory system, the FRC is determined by two equal and opposing forces: the elastic recoil pressure of the lung (the natural tendency of the lung to retract inward) and the outwardly directed chest wall recoil pressure. The interaction of these opposing forces is transmitted across the pleural space and accounts for the slightly subatmospheric or negative intrapleural pressure of −3 to −5 cm of water (H_2O) at FRC. Similarly, at total lung capacity (TLC), defined as the volume of gas contained in the lungs after maximal inspiration, the outward force of contraction of the inspiratory muscles is matched against the inward elastic recoil pressure of the lungs and chest wall. At residual volume (RV), which is the volume of gas remaining in the lungs after a maximal forced expiration, the muscles of expiration and the inward elastic recoil of the lung are counterbalanced by the outward elastic recoil of the chest wall.

The RV cannot be measured directly using a spirometer, nor can the FRC and TLC that contain RV. For the determination of FRC, from which RV and TLC are then extrapolated from spirometric measurements, the use of gas-dilution techniques (helium dilution, nitrogen washout) and whole-body plethysmography are employed.

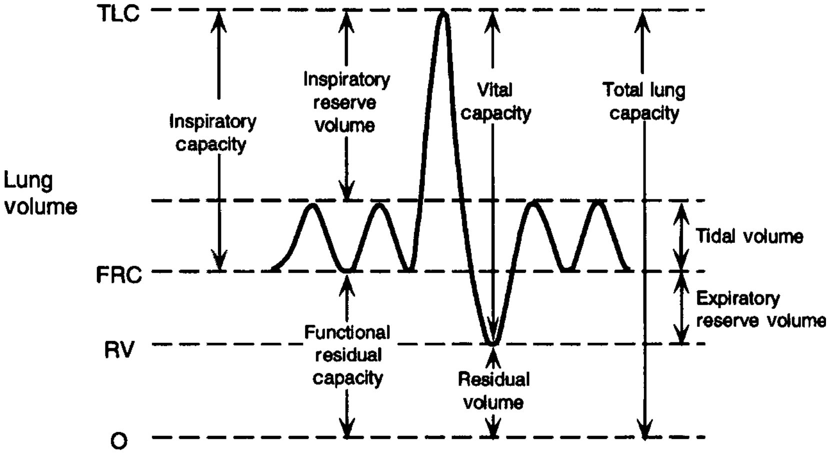


Figure 5–1 *Lung volumes. FRC, functional residual capacity; RV, residual volume; TLC, total lung capacity. (From Hlastala MP, Berger AJ. Physiology of Respiration. New York: Oxford University Press; 1996. Reprinted with permission.)*

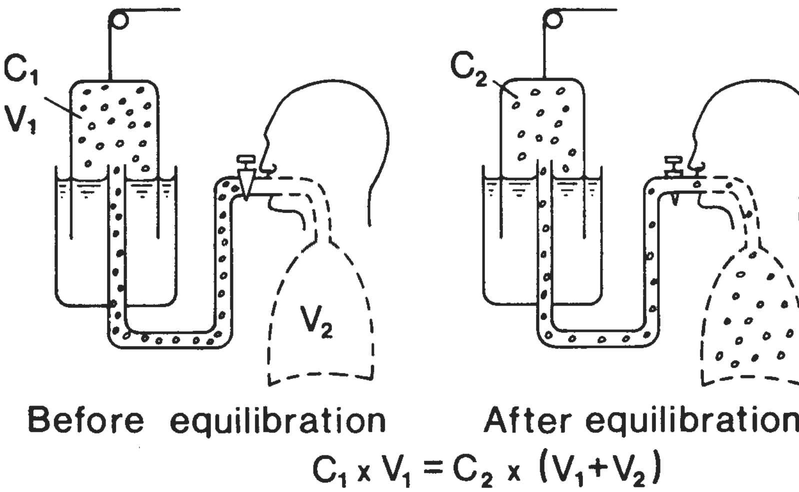


Figure 5–2 *The helium-dilution technique C_1, initial concentration; C_2, final concentration; V_1, known volume; V_2, final volume, which equals functional residual capacity. (From West JB. Respiratory Physiology—the Essentials. 5th ed. Baltimore: Williams & Wilkins; 1995. Reprinted with permission.)*

The helium-dilution technique (**Fig. 5–2**) exploits the concept that in a closed system, the total amount of a gas, defined as the product of its concentration and volume, remains constant. If a known initial concentration of a helium gas mixture (C_1) is contained in a spirometer circuit of known volume (V_1), and is then inspired by taking tidal breaths from FRC, over several minutes the helium gas mixture will equilibrate in a final volume that includes the subject's lungs. No helium is lost during testing due to its virtual insolubility in blood. The FRC can therefore be determined because $C_1 \times V_1 = C_2$ (the final measured concentration of the helium gas mixture after equilibration) $\times (V_1 + V_2)$, where V_2 equals the FRC.

The nitrogen-washout technique involves tidal inspiration of 100% oxygen through a one-way valve and collection of the expired gas in a spirometer. Over a period of minutes, the nitrogen contained in the lungs at baseline (assumed, like ambient air, to comprise ~80% of the initial lung volume) will be washed out, and its concentration entering expired air as detected by a nitrogen analyzer will approach zero. At this point, the total volume of expired gas is measured and multiplied by the final concentration of nitrogen in the expirate. The volume occupied by nitrogen in the subject's lungs at the outset can then be determined; because this represents roughly 80% of the initial lung volume, multiplying this value by 1.25 provides an estimate of FRC.

The measurement of FRC by whole-body plethysmography (**Fig. 5–3**) is based on Boyle's law, which states that under constant temperature, pressure times volume is a constant (K). The subject sits in an airtight box of known volume and makes inspiratory efforts from FRC while connected to a mouthpiece with a closed shutter. The inspiratory effort results in expansion of gas in the lungs associated with an increase in lung volume and fall in intrathoracic pressure (measured by pressure changes at the mouth). Similarly, in this closed system, there is a corresponding decrease in box volume of exact magnitude to the increase in lung volume (V) and an increase in box pressure. The lung volume can be determined by looking at changes in box pressure and volume: $P_1 \times V_1 = P_2 \times (V_1 - V)$, where P_1 and P_2 are box pressures before and after the inspiratory effort, and V_1 is the initial box volume. The FRC can be measured next by examining pressure and volume changes in the lungs: $P_3 \times V_2 = P_4 \times (V_2 + V)$, where P_3 and P_4 are mouth pressures before and after the inspiratory effort, and V_2 is the FRC.

Determinations of lung volume obtained by either the helium-dilution or nitrogen-washout technique measure gas in direct communication with the mouth

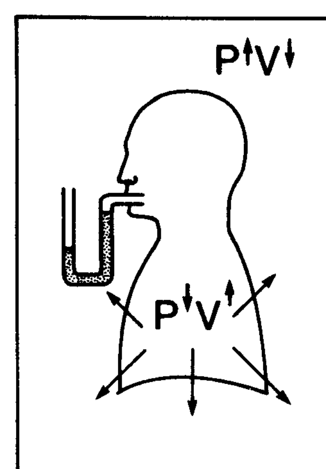


Figure 5–3 *Whole-body plethysmography. P, pressure; K, constant V, volume. (From West JB: Respiratory Physiology—the Essentials. 5th ed. Baltimore: Williams & Wilkins; 1995. Reprinted with permission.)*

but are unable to measure gas trapped behind obstructed airways. This is in contrast to whole-body plethysmography that measures the total volume of gas in the lungs. Gas-dilution techniques may therefore underestimate lung volumes in subjects with obstructive lung disease.

PULMONARY FUNCTION TESTS: PATTERNS OF IMPAIRMENT

Measurements of lung volume generally are considered in the normal range if they fall within 20% of a predicted value derived from population studies and based on subject age, gender, and height. During a forced expiratory maneuver, positive intrapleural pressure is transmitted to the airways, leading to rapid expulsion of gas from the lungs. In conditions characterized by narrowing of the airway lumen due to inspissated secretions and mucosal inflammation (as in chronic bronchitis or asthma), or by decreased lung elastic recoil and loss of radial traction on the airways (as in emphysema), forced expiration is associated with a marked increase in airway resistance and a corresponding drop in expiratory flow rates. Early airway closure can result in air trapping. As measured by spirometry, this translates into a decreased FEV-1 and decreased FEV-1:FVC ratio (<70%). The FVC is usually normal but may be decreased, the latter due to air trapping and a consequent rise in RV (especially in emphysema). In severe emphysema, the loss of lung elastic recoil, presumably due to elastin degradation and destruction of alveolar septa, leads to hyperinflation, a greater outward pull imparted by the recoil pressure of the chest wall at FRC, and a resultant increase in TLC.

Restrictive dysfunction on pulmonary function testing can result from parenchymal lung disease (e.g., pulmonary fibrosis, granulomatous lung disease); neuromuscular disease (e.g., amyotropic lateral sclerosis, multiple sclerosis, poliomyelitis); alveolar filling processes (e.g., pulmonary edema); or chest wall disease (e.g., severe kyphoscoliosis). A diminished TLC has been identified as the definitive measure of restrictive impairment, although a decreased vital capacity in the absence of significant air trapping is highly suggestive and is often used to follow the course of disease. The FEV-1 is usually diminished to the same extent as the FVC, and the FEV-1:FVC ratio consequently is normal.

Flow–Volume Loops

Flow–volume loops measure the overall mechanical properties of the lung. Instantaneous flow rates are recorded against lung volumes during maximum effort, thus evaluating the change in lung volume over time (see the normal flow–volume loop in **Fig. 5–4**, center). The upper limb of the flow–volume loop represents instantaneous forced expiratory flow rates beginning from a maximal expiratory maneuver from TLC. The uppermost point of the expiratory limb represents the peak expiratory flow rate (PEFR), which usually occurs at near 90% of vital capacity and approximates 8 L per second. Reduction in the PEFR and increased concavity of the expiratory limb (implying flow limitation in the smaller airways) are characteristic findings in subjects with obstructive lung disease (**Fig. 5–4**, bottom). The flow–volume loop in subjects with restrictive lung disease (**Fig. 5–4**, top) often appears as a miniaturized version of the predicted normal curve with a symmetric diminution in lung volumes. Interstitial/fibrotic lung

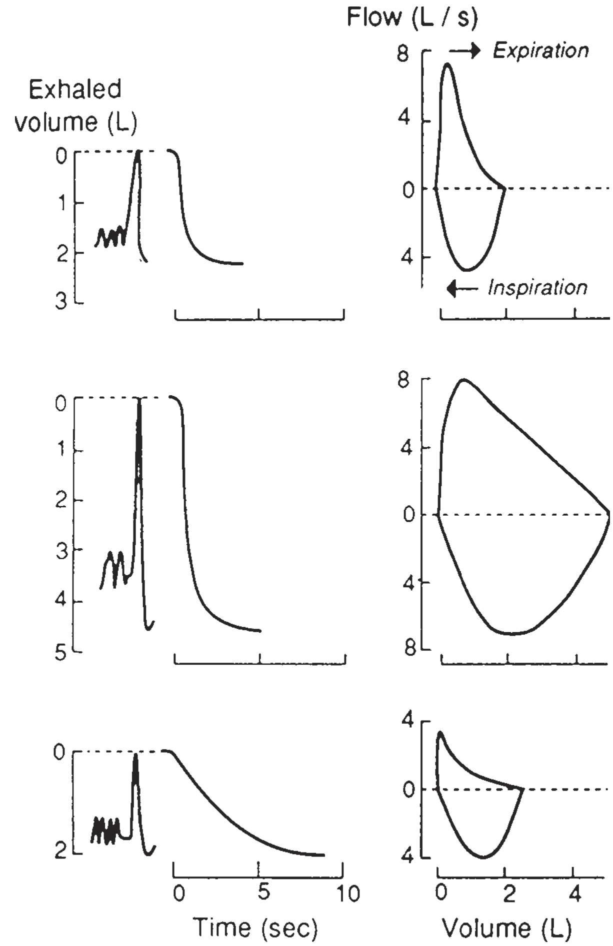


Figure 5–4 *Flow–volume loops. Normal expiratory and inspiratory loops (middle), restrictive pattern (top), and obstructive pattern (bottom). (From Hlastala MP, Berger AJ. Physiology of Respiration. New York: Oxford University Press; 1995. Reprinted with permission.)*

disease is characterized by less compliant (stiffer) lungs that resist full expansion, as well as increased elastic recoil pressure associated with greater expiratory flow rates than predicted at comparable lung volumes in normal subjects (isovolume comparisons).

Flow–volume loops in subjects with upper airway obstruction will be discussed in the section dealing with physiology of the extrathoracic airway and trachea.

MEASUREMENT OF VENTILATION

Resting Ventilation

In the upright lung, ventilation per unit lung volume is increased at the lung bases compared with the apices. This is primarily due to the effects of gravity because basal lung must exert a greater pressure to support the weight of more apical lung tissue. The result is a less negative intrapleural pressure at the lung base than at the apex, and basal alveoli that are less distended than apical alveoli at resting end-expiration (FRC). With inspiration, a greater volume of air is thus exchanged (greater ventilation) with the relatively less distended and more compliant alveoli at the lung base. As expected, ventilation to the upper lung zones increases and equals that in the lower lung zones with recumbency.

Minute ventilation is the volume of gas exhaled (VE) per minute, calculated by multiplying the TV by the number of breaths per minute (i.e., TV $\times$ respiratory rate). The measurement of minute ventilation plays little part in a routine pulmonary assessment. Normal values are difficult to obtain because, once the patient realizes that breathing is monitored, it is no longer an unconscious process. In the mechanically ventilated patient, however, a high VE ($>$10 L/minute) required to maintain acceptable gas exchange implies wasted ventilation and is one parameter among many that may be used as a weaning criterion.

Dead Space Ventilation

Expired minute ventilation (VE) in liters per minute is the sum of the alveolar ventilation (VA), the volume of gas that is exchanged between blood and alveoli, and the volume of gas wasted ventilating the dead space (VD). The physiological dead space consists of the anatomical dead space, or that segment of the tracheobronchial tree containing no alveolar gas exchange units (the conducting zone), plus the alveolar dead space, the volume of gas entering unperfused alveoli [alveoli with infinite ventilation/perfusion (V/Q) ratios]. Assuming a TV of ~500 mL in a normal healthy individual weighing 150 pounds, ~350 mL per breath will participate in gas exchange, and the dead space per breath will comprise 150 mL. Because, in this healthy individual, alveolar dead space is minimal, the physiological dead space approximates the anatomical dead space (150 cc). Conversely, a low cardiac output state can result in derecruitment and underperfusion of pulmonary capillaries and a marked increase in alveolar dead space ventilation (alveolar capillary units with high V/Q ratios). In this example, the physiological dead space can greatly exceed the anatomical dead space due to the addition of alveolar dead space. Other conditions associated with an increase in alveolar dead space include pulmonary embolism (occlusion of pulmonary blood flow), hemorrhage (decreased right ventricular output with pulmonary capillary derecruitment), and positive pressure ventilation plus positive end-expiratory pressure (increased alveolar pressure with compression of pulmonary capillaries).

The anatomical dead space can be estimated from body weight, ~1 mL per pound, or it can be measured directly using a single-breath technique (Fowler's method). With this technique, the anatomical dead space is the volume of gas contained in the airways whose composition remains unchanged from that of inspired air. It has been shown that the extrathoracic pharynx and mouth contribute ~66 mL of the anatomical dead space to the average 150 mL of a normal-sized man. The anatomical dead space is decreased by ~60% following tracheostomy.

The physiological dead space is measured using Bohr's equation: $VD/VT = [(PACO_2 - PECO_2)/PACO_2]$, where VD/VT is the dead space to tidal volume ratio, $PACO_2$ the alveolar carbon dioxide tension, and $PECO_2$ the carbon dioxide tension of expired air. Bohr's method measures the volume of the lung that does not eliminate CO_2. Stated another way, any CO_2 collected in expired air must have come from alveolar-capillary units that participated in gas exchange, because inspired air contains negligible concentrations of CO_2. Therefore, the greater the difference per breath between $PACO_2$ and $PECO_2$, the greater the physiological dead space. Under normal conditions, the $PaCO_2$ is substituted for $PACO_2$ in the Bohr equation because these values are virtually equal. The normal range for the VD/VT is 0.2 to 0.35 during quiet breathing.

PULMONARY BLOOD FLOW AND INTRAPULMONARY SHUNT

In the upright subject, pulmonary blood flow decreases in a nearly linear fashion from the lung bases to the apices due to the decreasing hydrostatic pressures imposed by gravity. When lying supine, perfusion to the lung apices increases such that the distribution of blood flow from top to bottom becomes uniform, although flow to dependent (posterior) regions increases.

The upright lung hypothetically can be divided into three regions: in zone 1 (the lung apex), alveolar pressure can, in certain disease states, exceed pulmonary artery pressure, leading to capillary collapse and cessation of flow (alveolar dead space). In zone 2 (midlung region), pulmonary artery pressure exceeds alveolar pressure, although pulmonary venous pressure is still low and less than alveolar pressure. In this circumstance, blood flow is determined by the arterial-alveolar pressure difference, and flow is limited at the point of collapse of the pulmonary vein. The resulting hemodynamics, sometimes referred to as the Starling resistor or waterfall effect, will lead to venous compression, recruitment of pulmonary capillaries upstream from the point of obstruction, and increases in pulmonary hydrostatic pressure to capillary beds lower in the zone. In zone 3 (at the lung bases), pulmonary arterial pressure is greater than pulmonary venous pressure, which in turn exceeds alveolar pressure. Blood flow is thus determined by the arterial-venous pressure difference. Capillary distention in the face of relatively high hydrostatic pressures, rather than recruitment, is the chief mechanism underlying increased flow in zone 3.

Intrapulmonary shunt refers to perfusion of lung units that are not ventilated (V/Q ratio of zero). End-capillary blood perfusing a shunted lung unit will therefore have a gas composition mirroring mixed venous blood (partial oxygen pressure [PO_2] 40 mm Hg, partial CO_2 pressure [PCO_2] 45 mm Hg). This contrasts with alveolar dead space (V/Q ratio of infinity), where the composition of end-capillary blood resembles that of inspired gas (PO_2 150 mm Hg, PCO_2 0 mm Hg) because there is no contribution from mixed venous precapillary blood. An important feature of shunt physiology is that administration of 100% inspired oxygen is unable to correct hypoxemia because obstructed alveolar units remain unventilated; the percentage of shunted blood to total blood flow in the lungs, known as the shunt fraction, remains unchanged. The determination of the shunt fraction requires insertion of a pulmonary artery catheter for measurement of arterial and mixed venous oxygen contents, and estimation of end-capillary oxygen content from the alveolar PO_2 and the oxyhemoglobin dissociation curve (see Gas Exchange below).

Another important adaptation of the pulmonary circulation is termed hypoxic vasoconstriction. In conditions associated with a drop in alveolar PO_2 (and/or lowering of blood pH), contraction of smooth muscle in the walls of small arterioles effectively limits perfusion and directs blood flow away from hypoxic lung regions. This serves to facilitate matching of ventilation to perfusion, thereby minimizing effects on gas exchange.

VENTILATION/PERFUSION (V/Q) RELATIONSHIPS

Of the five etiologies for hypoxemia (hypoventilation, decreased fractional inspired oxygen concentration, diffusion abnormality, shunt physiology, and ventilation–perfusion inequality), ventilation–perfusion inequality is the most common. As was previously discussed for the normal upright lung, both ventilation and perfusion increase when going from the apex to the base of the lung, although perfusion increases at a faster rate. Lung units at the apex therefore have high V/Q ratios, contributing to alveolar dead space, whereas units at the base have low V/Q ratios, exhibiting shunt physiology. Uneven matching of ventilation to perfusion, caused by multiple etiologies, can result in varying concentrations of gas in end-capillary blood. In severe cases, the end result is not only hypoxemia but also CO_2 retention. In milder cases, however, CO_2 retention is minimized by compensatory hyperventilation at the cost of increased respiratory work and wasted ventilation.

The alveolar-arterial (A-a) gradient provides an indirect assessment of V/Q inequality. The $PAO_2 = [(PB - 47 \text{ mm Hg}) \times FiO_2 - PCO_2/R]$, where PAO_2 represents alveolar oxygen tension, PB the barometric pressure (47 mm Hg is subtracted from PB to account for water vapor pressure in the airways), FiO_2 the fractional concentration of oxygen in air, and R the respiratory exchange ratio (ratio of CO_2 production to oxygen consumption in tissues in the steady state, usually around 0.8). If a subject is breathing room air ($FiO_2 = 0.21$) at sea level ($PB = 760$ mm Hg), with a normal PCO_2 of 40 mm Hg and an R value of 0.8, the PAO_2 approximates 100 mm Hg. The A-a gradient is determined by subtracting the arterial oxygen tension from PAO_2. A normal gradient is ~12 mm Hg, although this value increases with age and reflects increased V/Q inequality over time.

PULMONARY MECHANICS

Air flows from a region of high pressure to low pressure. In order for inspiration to take place, alveolar pressure must be overcome by atmospheric pressure. Expansion of the chest wall involves contraction of the muscles of inspiration that pull on the lateral wall of the thorax,

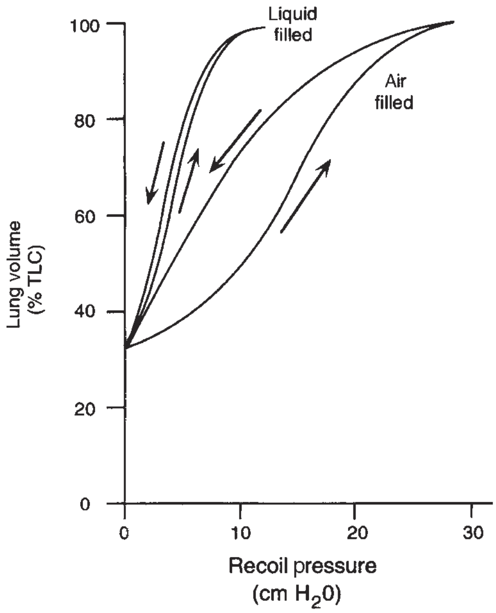


Figure 5–5 *Static pressure–volume of the lung. TLC, total lung capacity. (From Hlastala MP, Berger AJ. Physiology of Respiration. NewYork: Oxford University Press; 1996. Reprinted with permission.)*

which in turn expands the lung, alveoli, and the smaller airways. Alveolar distention leads to a decrease in alveolar pressure, thereby permitting the flow of inspired air via the mouth, nose, and trachea to the intrathoracic airway.

Lung compliance (V/P) is defined as the change in lung volume observed per unit change in the pressure gradient between the intrapleural space and alveoli (the transpulmonary pressure). A static pressure–volume curve of the lung (**Fig. 5–5**) can be generated by interrupting a full expiratory or inspiratory maneuver in steps of ~500 cc, and after a few seconds of equilibration, measuring the corresponding intrapleural pressure. Esophageal pressure is used as a surrogate measure of intrapleural pressure and involves swallowing a small balloon-tipped catheter to the level of the midesophagus. At periods of no flow (during breath holding), the glottis should remain open to ensure that alveolar pressure equals zero (equivalent to atmospheric pressure). Static compliance is represented as the midcurve slope of the pressure–volume curve, by convention derived from the deflation curve. A reflection of the elastic properties of the lung, static compliance is increased (curve shifted upward and to the left) with advancing age and in subjects with emphysema. It is decreased (shifted downward to the right) in association with interstitial/fibrotic lung disease, and alveolar filling processes (e.g., cardiogenic pulmonary edema, adult respiratory distress syndrome, pneumonia).

Note in **Fig. 5–5** that at any given transpulmonary pressure, there is a correspondingly higher volume on the deflation as opposed to inflation curve. This phenomenon is known as hysteresis, and is attributed to pulmonary surfactant, a complex phospholipid moiety produced by type II alveolar lining cells. By decreasing the surface tension of alveolar lining fluid, surfactant serves many functions that include stabilizing alveoli from collapse, increasing lung compliance, preventing alveolar edema by reducing interstitial hydrostatic pressure, and decreasing the respiratory work required to expand alveoli.

Dynamic compliance differs from static compliance in that the pressure–volume curves are generated during airflow, rather than by interruption of airflow at various lung volumes. This assumes importance when considering the effects of an increased respiratory rate. In normal subjects, when breathing frequency is increased to 60 to 90 breaths per minute, dynamic and static compliances are nearly equal. The same scenario in subjects with obstructive lung disease and increased airway resistance is associated with a fall in dynamic compliance compared with static compliance. Targeted lung units facing a high airway resistance require a longer time to fill and empty. Stated another way, their time constant (compliance × resistance) is prolonged. As the time for inspiration and expiration diminishes with increasing respiratory rates, lung units with long time constants may empty incompletely or may still be filling during expiration. Consequently, a smaller proportion of each subsequent tidal volume will distribute to these partially obstructed lung units. This reduces the effective lung volume that receives tidal breaths, resulting in a fall in compliance.

The work of breathing (pressure × volume) corresponds to the area subtending a dynamic pressure–volume curve and is the sum of work performed to overcome elastic forces and viscous forces (airway resistance + tissue resistance). Tissue resistance, normally accounting for a small percentage of the total respiratory work performed, refers to the resistance imparted by tissues as they slide over each other during lung and chest wall excursion. More importantly, airway resistance (RAW) can account for large increases in respiratory work in subjects with obstructive lung disease or upper airway obstruction. Defined as the pressure difference applied between the alveoli and the mouth per unit of airflow, RAW is routinely measured by whole-body plethysmography that measures changes in pressure and flow at the mouth during respiratory efforts. Under normal conditions, the chief site of RAW resides in the medium-sized bronchi, although the upper airway contributes a major share, particularly when nose breathing is employed.

GAS EXCHANGE

Oxygen Uptake and Delivery

Aerobic metabolism is completely dependent on oxygen (O_2) uptake from inspired air and its delivery to tissues. The binding of O_2 to hemoglobin (Hb) provides the major vehicle for O_2 transport by the blood. Under normal conditions, each gram of hemoglobin binds 1.39 mL of O_2. Because a normal Hb concentration is ~15 g/100 mL blood, and assuming an arterial O_2 saturation (SaO_2) of 100% (the percent saturation of Hb with O_2), the arterial O_2 carrying capacity of Hb is 1.39 mL O_2/g Hb × 15 g Hb/100 mL blood × SaO_2/100, or 20.8 mL O_2/100 mL blood. Oxygen dissolved in the blood, conversely, contributes little to O_2 transport. According to Henry's law, the amount of dissolved O_2 is proportional to its partial pressure; for an arterial PO_2 of 100 mm Hg, only 0.3 mL of O_2 is dissolved in each 100 mL of blood. In this example, the total volume of O_2 carried in arterial blood, termed the arterial oxygen content (CaO_2), is 20.8 mL O_2 (carried by hemoglobin) plus 0.3 mL O_2 (dissolved O_2), or 21.1 mL O_2/100 mL blood. Mixed venous blood in the normal resting state has a PO_2 of 40 mm Hg and a mixed venous O_2 saturation (SvO_2) of 75%. Assuming, once again, an Hb concentration of 15 g/100 mL blood, this corresponds to a mixed venous oxygen content (CvO_2) of 15.8 mL O_2/100 mL blood, a value three quarters of the CaO_2. If the rate that O_2 is delivered to tissues, termed the oxygen delivery, is defined as $CaO_2 \times Qt$ (the cardiac output), and oxygen consumption is defined as $(CaO_2 - CvO_2) \times Qt$, only 25% of the oxygen delivered in the normal resting state is consumed at the tissue level. Ample reserve therefore exists for augmented O_2 uptake and utilization during times of stress or exercise.

Oxygen binds reversibly with Hb to form oxyhemoglobin: $O_2 + Hb = HbO_2$. The oxyhemoglobin dissociation curve (**Fig. 5–6**) defines the relationship between oxygen saturation and the partial pressure of oxygen in blood. The sigmoid-shaped curve reflects the allosteric binding properties of O_2 with Hb. Each Hb molecule has the capacity to bind four molecules of O_2 such that the binding of the first few molecules of O_2 stimulates more favorable binding of additional O_2 molecules. The asymptotic upper portion of the curve demonstrates that oxygen saturation will remain above 90% until the PO_2 falls below 60 mm Hg. Consequently, marked drops in alveolar PO_2 have little effect on oxygen loading in the lungs.

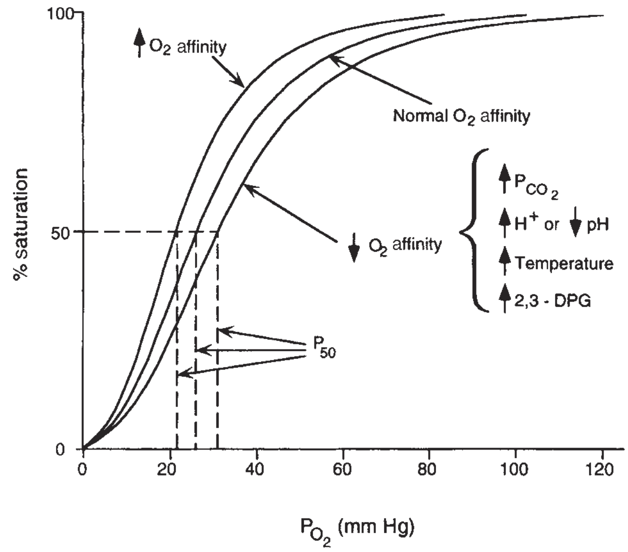


Figure 5–6 *Oxygen (O_2) dissociation curve. Physical factors that cause a shift in the oxyhemoglobin curve include increased partial pressure of carbon dioxide (CO_2), decreased pH, increased temperature, and increased concentration of 2,3-diphosphoglycerate (2,3-DPG). (From Hlastala MP, Berger AJ. Physiology of Respiration. New York: Oxford University Press; 1996. Reprinted with permission.)*

The steep portion of the oxyhemoglobin dissociation curve shows that small decrements in capillary PO_2 are associated with rapid falls in oxygen saturation. Thus oxygen unloading at the tissue level is facilitated in response to minor drops in PO_2. The P50, which is the PO_2 corresponding to an oxygen saturation of 50%, is a useful measure for examining hemoglobin's O_2 affinity. A rightward shift in the curve (associated with an increased P50) signifies that more O_2 has been unloaded at the tissue level for any given PO_2. Physical factors that cause a rightward shift include acidemia, hypercarbia, fever, and 2,3-diphosphoglycerate (2,3-DPG) (**Fig. 5–6**). The 2,3-DPG, which is produced by erythrocytes during anaerobic glycolysis and elevated in red blood cells during chronic hypoxia, is associated with decreased affinity of Hb for O_2. Of interest, blood stored for as little as 1 week has been shown to have low levels of 2,3-DPG. Use of banked blood may therefore impair O_2 unloading, unless normal levels of 2,3-DPG are restored.

Diffusing Capacity

The diffusion of gas across tissue as defined by Fick's law states that the volume of a gas transferred is proportional to the surface area and the partial pressure difference of the gas across the interface, and inversely proportional to the thickness of the tissue sheet.

The large surface area and ultrathin width of the alveolar-capillary membrane therefore constitute an ideal environment for gaseous diffusion.

Fick's law also states that diffusion is proportional to the diffusivity (diffusion constant) of the gas, which is the solubility of the gas divided by the square root of the molecular weight. Diffusion of carbon dioxide is ~20 times that of oxygen due to its higher solubility in blood, and thus its transfer across the alveolar-capillary interface is usually maintained even in the presence of extensive lung disease. Diffusion of oxygen, alternatively, may be limited in lung diseases characterized by a thickened (interstitial lung disease) or destroyed (emphysema) alveolar-capillary interface, or by intra-alveolar edema [acute respiratory distress syndrome (ARDS), congestive heart failure].

Carbon monoxide (CO), rather than oxygen, is routinely used to measure the diffusing capacity of the lung (DLCO). The single-breath method involves inhalation of a small concentration of CO, a 10-second breath hold, and measurement of CO in an expired alveolar gas sample (the less CO recovered in the expired sample, the greater the diffusion across the alveolus into capillary blood). Similar to, but with an avidity 200 times that of oxygen, CO binds to hemoglobin. Because hemoglobin provides a veritable sink for CO binding, the partial pressure of CO in pulmonary capillary blood remains negligible, and the pressure gradient from alveolus to red cell is maintained. Because binding with hemoglobin is an inherent component of the diffusing capacity measurement, adjustments should be made for hemoglobin concentration. Other factors associated with a high DLCO include alveolar hemorrhage (more hemoglobin-binding sites), high altitude (the lower ambient oxygen tension leads to a lower mixed venous oxygen saturation and more available sites for CO binding), and recumbency (increased pulmonary capillary blood volume due to increased perfusion to the upper lung zones).

PHYSIOLOGY OF THE TRACHEA

Diseases commonly encountered by the otolaryngologist include those affecting the upper airway and trachea. In **Fig. 5–7**, representative flow–volume loops depict the patterns of upper airway and tracheal obstruction that may be seen in these conditions. To better understand the pathophysiology of fixed or variable extra- and intrathoracic obstruction, it is necessary to review the physiology of the upper airway and trachea.

The extrathoracic airway comprises the naso- and oropharynx, larynx, and proximal trachea to the level of the sternal notch. Congenital or acquired anatomical abnormalities, infections, and tumors involving upper airway structures can be associated with variable degrees of airflow obstruction. Some of the more common etiologies are nasal polyps, adenoidal or tonsillar hypertrophy (especially in children), acute epiglottitis, and nasopharyngeal or laryngeal tumors. Upper airway collapse during sleep, usually accompanied by loud snoring, is characteristic of subjects with the obstructive sleep apnea syndrome (OSAS) who can have predisposing anatomical abnormalities such as macroglossia, retrognathia, tonsillar hypertrophy, redundant pharyngeal soft tissue, and chronic nasal congestion and/or polyps. Tracheal obstruction can be the result of benign or malignant tumors either extrinsically compressing the lumen or protruding endotracheally. Tracheomalacia and tracheal stenosis can lead to upper airway obstruction as a complication of prior tracheostomy. A large goiter is a rare cause of tracheal obstruction.

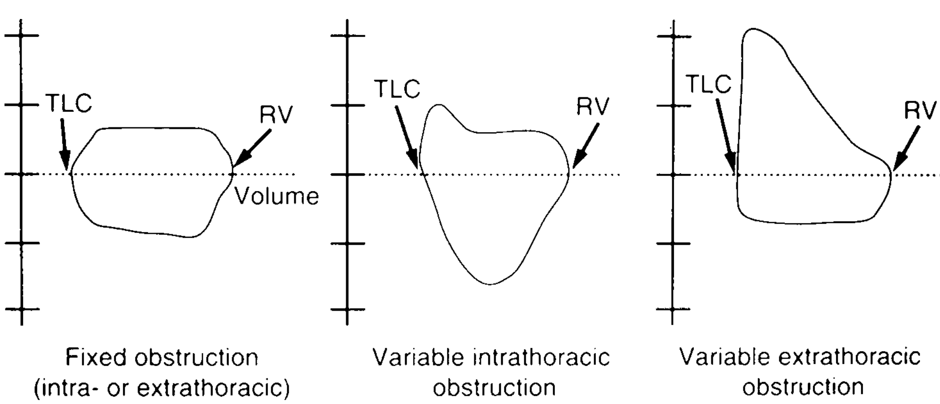


Figure 5–7 *Flow-volume loops in airway obstruction. RV, residual volume; TLC, total lung capacity. (From Hlastala MP, Berger AJ. Physiology of Respiration. New York: Oxford University Press; 1996. Reprinted with permission.)*

The degree of airway compression is related to the external compressive pressure on the tracheal wall and the compressibility of the airway. This phenomenon is known as dynamic compression. Because the cartilaginous walls of the trachea are normally rigid, only the membranous components can shift during air movement. Indeed, collapse of the membranous portion can lead to transient obliteration of the tracheal lumen. This serves an adaptive physiological function for force generation during expectoration. Flaccidity of tracheal segments, however, may result in variable grades of upper airway obstruction.

In cases of extrathoracic airway obstruction, inspiratory stridor is due to generation of subatmospheric intraluminal pressures during inspiration. The transmural pressure imparted on the extrathoracic airways depends on several factors: airway opening pressure, the resistive pressure drop during inspiration, and the pressure drop associated with the convective acceleration of gas (known as the Bernoulli effect). Factors leading to increased ventilatory demands and high inspiratory flow rates that exacerbate extrathoracic airway compression include anxiety, tachypnea, fever, and any process that increases the work of breathing. The negative transmural pressure (intraluminal−atmospheric pressure) and resultant airway compression during inspiration will be limited to the extrathoracic airway because the transpulmonary pressure gradient from the alveolus to the intrapleural becomes more negative with inspiration, dictating against intrathoracic obstruction.

The tracheal lumen changes dimensions depending on the phase of respiration. Variable external pressure gradients act on the cervical and intrathoracic trachea. As discussed previously, during quiet respiration the intrathoracic pressure is negative relative to atmospheric pressure. It fluctuates between −7 cm H_2O and −2 cm H_2O during inspiration and expiration, respectively. During a Valsalva maneuver there is an increase in the diameter of the cervical trachea, while the transmural pressure gradient of zero maintains a stable intrathoracic tracheal lumen.

Flaccidity of tracheal segments may result in variable grades of upper airway obstruction. Extrathoracic tracheomalacia gives rise to inspiratory obstruction, and intrathoracic tracheal obstruction leads to expiratory collapse. The patency of the intrathoracic trachea is supported by the surrounding negative intrapleural pressure (except during forced expiration). In the cervical trachea, the force maintaining patency with each inspiration is the cricoid attachment superiorly.

Flow−volume curves are a useful way to assess extrathoracic or tracheal obstruction. A fixed obstruction (intra- or extrathoracic) affects both inspiration and expiration, as seen in **Fig. 5−7** (left). Due to stiffness of the airway, the transmural pressure gradient has no effect on the airway, and there is a truncation of both phases of the flow−volume loop. It is not possible to determine if the obstruction is intra- or extrathoracic. Bronchoscopy is valuable in determining the presence of foreign bodies or scarring.

During forced expiration in the setting of variable intrathoracic obstruction, there is a decrease in the cross-sectional area of the airway and decreased forced expiratory flow, as shown in **Fig. 5−7** (center). There is no change in the inspiratory loop under these conditions due to an increase in the cross-sectional area of the intrathoracic trachea following generation of a large negative intrapleural pressure. Variable intrathoracic obstruction is seen in tumors affecting the trachea.

Fig. 5−7 (right) shows that in the setting of variable extrathoracic obstruction, there is a decrease in the cross-sectional area of the upper airway, signified by a truncated inspiratory loop. During a forced expiration, the cross-sectional area of the airway in the setting of variable extrathoracic obstruction increases due to an increase in the intraluminal pressure. Therefore, the expiratory phase of the flow−volume curve is essentially normal. Variable extrathoracic obstruction can be secondary to tumors, weakened pharyngeal muscles (e.g., OSAS), paralyzed vocal cords, lymphadenopathy, fat deposits, or inflammation.

MECHANICAL VENTILATION

In the last century, during the polio epidemic, mechanical ventilation first achieved widespread use when negative-pressure ventilators, so-called iron lungs, were used to assist the failing muscles of respiration. Today, the routine use of positive-pressure mechanical ventilation underscores the intricate interplay between the patient's lung mechanics and the settings imposed by the ventilator. The objectives of mechanical ventilation are to improve pulmonary gas exchange, relieve respiratory distress, alter pressure−volume relations, and permit lung and airway healing. However, inherent risks are involved; oxygen toxicity, endotracheal tube complications (laryngeal injury, subglottal or tracheal stenosis, tracheomalacia, sinusitis), volume-induced alveolar injury (volutrauma), barotrauma, decreased cardiac output, pneumonia, and psychological problems are well-known complications of mechanical ventilation.

Common Modes of Mechanical Ventilation

Continuous Mandatory Ventilation and Assist Control Ventilation

Continuous mandatory ventilation (CMV) has been supplanted by assist control (AC) ventilation. In CMV, the ventilator is set to deliver a fixed, nontriggered tidal volume at a fixed respiratory rate, ignoring the patient's ventilatory drive. Assist control was designed to eliminate the synchronization problems inherent in CMV. During a patient-triggered inspiratory effort, there is a pressure drop in the airway sensed by the ventilator followed by delivery of a preset tidal volume. The ventilator is also preset to deliver a minimum number of breaths per minute, analogous to CMV, should the patient fail to trigger the ventilator. The work of breathing is substantial for patients receiving AC but may be reduced by proper sedation or the use of neuromuscular blocking agents in severe cases.

Intermittent Mandatory Ventilation and Synchronized Intermittent Mandatory Ventilation

Intermittent mandatory ventilation (IMV) is a combination of CMV and spontaneous respiratory efforts. The ventilator is set to deliver nontriggered breaths at a fixed volume and rate. Patient-triggered breaths are also possible similar to the AC mode, but rather than delivering a preset TV, the TV generated is derived entirely from patient effort. Thus total minute ventilation is equal to the sum of tidal volumes generated by spontaneous efforts plus machine-delivered breaths. Synchronized intermittent mandatory ventilation (SIMV) is a variation of IMV in which the ventilator senses spontaneous efforts and times its own mechanical breaths to prevent breath stacking. Like AC, SIMV imposes a significant respiratory and cardiac workload.

Pressure Support Ventilation

Pressure-support ventilation (PSV) differs from AC ventilation and IMV in that the physician sets a level of pressure (rather than volume) to augment spontaneous breathing efforts. Tidal volume is determined by the level of set pressure, patient effort, and pulmonary mechanics (less compliant lungs will generate a smaller tidal volume for a given level of pressure support and vice versa). At any constant level of patient effort, the greater the pressure support, the greater the combined volume. At any constant level of pressure support, the greater the inspiratory effort, the greater the combined volume. Advantages of PSV include patient comfort due to ventilator synchronization with spontaneous respiratory efforts and decreased work of breathing. Pressure support ventilation, sometimes in conjunction with SIMV, is commonly used as a weaning mode.

Positive End-Expiratory Pressure

Positive end-expiratory pressure (PEEP) is used as an adjunct to the various ventilator modes thus far described. During normal breathing, airway pressure is zero (atmospheric) at end-expiration. If a subject exhales into tubing, the distal end of which is submerged under a column of water, the end-expiratory pressure becomes positive relative to the atmosphere. This applied PEEP is also known as extrinsic PEEP. PEEP can also be measured in the spontaneously breathing subject, especially in someone with chronic obstructive pulmonary disease (COPD), high minute ventilation, and gas trapping. This phenomenon is termed auto-PEEP (or occult PEEP). Similarly, extrinsic PEEP can be applied in the mechanically ventilated patient who exhales against a preset pressure delivered by a one-way valve added to the expiratory circuit. Auto-PEEP, an unwanted complication of positive-pressure ventilation, can result in high airway pressures and an increased risk of barotrauma, and commonly occurs in the setting of bronchospasm, or expiratory airflow obstruction from any cause.

Extrinsic PEEP dialed into the ventilator circuit is useful for patients with alveolar filling processes such as the ARDS. By redistributing lung water from the alveoli to the perivascular interstitial space, intrapulmonary shunting is reduced, leading to an increase in arterial oxygen tension. Addition of PEEP can influence lung mechanics by elevating FRC, shifting tidal breathing to a more compliant portion of the pressure–volume curve, and reducing the work of breathing. In acute lung injury, recent evidence suggests that PEEP may promote the healing process by maintaining the patency of alveoli throughout the respiratory cycle, rather than subjecting alveoli to repeated cycles of opening and closing with each mechanically driven positive pressure breath.

Continuous Positive Airway Pressure

Continuous positive airway pressure (CPAP) was devised to overcome the problem of increased inspiratory work. Here, the positive pressure is not confined to expiration but is sustained throughout the breathing cycle. For some patients with severe airflow obstruction, the addition of CPAP may cause tidal volume to increase as pressure support or the natural force of breathing effort becomes more effective. For stronger patients, CPAP may be used as a counterspring against which expiratory muscles can store energy for release during subsequent inspiration.

In a child with upper respiratory tract infection, airway congestion, fever, and anxiety may begin a cycle of airway compression, decreased airway radius, and increased work of breathing. In some extreme cases, simple maneuvers such as nasal drops and postural drainage may not be sufficient to overcome the airway obstruction, and positive airway pressure (CPAP) may be needed to alleviate the extrathoracic obstruction.

Suggested Readings

Hlastala MP, Berger AJ. Physiology of Respiration. New York: Oxford University Press; 1996

Tobin MJ. Mechanical ventilation. N Engl J Med 1994;330(15): 1056–1061

West JB. Respiratory Physiology—the Essentials. 5th ed. Baltimore: Williams & Wilkins; 1995

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. Which of the following lung capacities cannot be measured by a spirometer?
   A. Total lung capacity
   B. Inspiratory capacity
   C. Vital capacity
   D. Functional residual capacity
2. If arterial oxygen content is 19 mL/100 mL of blood, and mixed venous oxygen content is 16 mL/100 mL of blood, what blood flow is necessary to sustain a volume of oxygen (VO_2) of 400 mL/minute?
   A. 12 L/minute
   B. 13.33 L/minute
   C. 10 L/minute
   D. Cannot measure with above information
3. Which of the following in arterial blood has the most important control on normal ventilation conditions?
   A. PCO_2
   B. pH
   C. PO_2

Chapter 6

BIOLOGY AND TREATMENT OF SLEEP APNEA

HECTOR P. RODRIGUEZ AND DIANA V.-A. BERGGREN

In 1836, Charles Dickens wrote *Posthumous Papers of the Pickwick Club*, in which he described the classical pickwickian syndrome in the character Joe, the "fat boy." In Dickens's novel, the reader is introduced to Joe as follows: "And on the box sat a fat and red-faced boy, in a state of somnolence." The description is pretty much on target in describing the alveolar hypoventilation syndrome, an extreme form in the spectrum of sleep disordered breathing (SDB) syndromes.

Approximately 50 million Americans snore, and 20 million Americans suffer from a sleep apnea syndrome. Together these syndromes are responsible for increases in spousal complaints and more importantly carry an increased risk of cardiovascular disease and premature death.

Under the label of SDB are classified a spectrum of disease processes, from snoring to the alveolar hypoventilation syndrome (**Fig. 6–1**), also known as the pickwickian syndrome. Obstructive sleep apnea syndrome (OSAS) lies somewhere at the center of the spectrum. Of concern is that snoring will progress in a significant number of patients into a more serious entity within the spectrum of SDB syndromes. When the disease entity reaches the levels of severe sleep apnea, it is associated with a 10% increase in mortality, usually related to the cardiovascular system (e.g., stroke, myocardial infarction, and deadly arrythmia). Severe sleep apnea syndrome also is strongly associated with increased body mass index (BMI, in kg/m^2), and it is an ongoing debate how much of the morbidity found in patients with OSAS is dependent on the sleep apnea syndrome per se respective to the increased BMI.

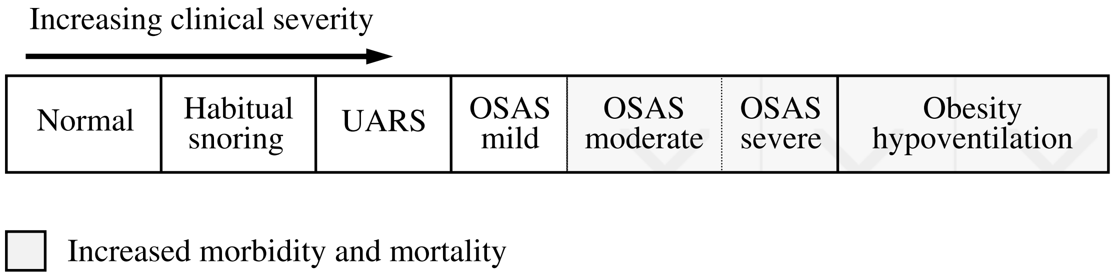


Figure 6–1 *Sleep disordered breathing syndromes. OSAS, obstructive sleep apnea syndrome; UARS, upper airway resistance syndrome.*

The U.S. Department of Transportation estimates that 200,000 automobile accidents yearly are sleep related, causing ~1500 road deaths per year. The loss of productivity in the work place as well as accidents while using heavy machinery at work in patients with sleep disorders has been documented in both the European and American literature.

The complexity of sleep apnea syndromes dictates that multiple specialties are involved in their treatment. Medical specialists such as general internists, endocrinologists, nutritionists, cardiologists, psychiatrists, and neurologists are often involved. On the surgical side, otolaryngologists, prostodontists, maxillofacial surgeons, and general surgeons may be involved.

PATHOGENESIS AND RISK FACTORS

Snoring is caused by the turbulence of air currents secondary to the partial collapse of the airway. Excessive relaxation of the pharyngeal musculature during sleep may be a single factor that causes the soft tissues of the soft palate, lateral pharyngeal walls, posterior pharyngeal wall, and tongue base to vibrate against each other. However, there are usually other anatomical factors; that is, tonsillar hypertrophy, lingual tonsil hypertrophy, nasal obstruction, and hypoplastic mandibles, which further contribute to the narrowing of the air space. Patients who are mouth breathers secondary to nasal obstruction may have excessive relaxation of the genioglossus muscle that retrodisplaces the tongue, reducing the size of the posterior airway space, which when combined with pharyngeal muscle relaxation, can cause snoring.

An increase in BMI of greater than 120% of ideal body weight and an increase in neck size (larger than 17 inches in males and 15 inches in females) correlate positively with the diagnosis of sleep apnea in greater than 60% of the patients diagnosed with this syndrome. Furthermore, according to several authors, maintaining or reducing weight regardless of the other modalities of therapy is the best predictor of overall success in the management of patients with OSAS. Patients with endocrine disorders like hypothyroidism and acromegaly are also at increased risk of developing the syndrome.

The degree of collapse, partial or total, of the pharyngeal musculature will determine whether the patient will suffer from snoring or obstructive apnea. Extreme fatigue, alcohol, sedative pills, and deeper stages of sleep diminish the pharyngeal muscle tone and can further aggravate the symptoms of snoring and sleep apnea. One pathophysiological mechanism underlying the progression from snoring to OSAS is believed to be local neurogenic lesions in the oropharynx caused by the low-frequency vibration of habitual snoring. This suggestion is based on results from biopsies of the palatopharyngeal muscle in which morphological abnormalities, including neurogenic signs, were found in snorers with and without OSAS. Together these data suggest that a disturbance in the efferent and/or afferent nerve pathways involved in the reflexogenic mechanism of the upper airway contribute to the pharyngeal collapsibility seen in patients with OSAS.

In patients with sleep apnea, the redundant tissues in the palate and retropharynx can be further elongated by the negative pressure effects of the lungs against a closed pharynx, creating a vicious cycle that can allow the progression of pathological events in the chain of sleep disordered breathing.

There are three key pathophysiological features of OSAS that can adversely affect the cardiovascular system: generation of exaggerated negative intrathoracic pressure, development of asphyxia during apnea, and arousal from sleep at the termination of apnea. These effects are schematically represented in **Fig. 6–2**. Muscle sympathetic nerve activity increases during apnea. It is also constantly elevated during the daytime in sleep apnea subjects. The heart rate decreases during apnea and increases afterward. The blood pressure has a typical pattern of variation during OSAS. It increases during the latter part of apnea with a peak immediately after apnea termination, whereafter the pressure declines. Simultaneously with arterial pressure, pulmonary artery pressure, central venous pressure, and intracranial pressure increase during apnea and

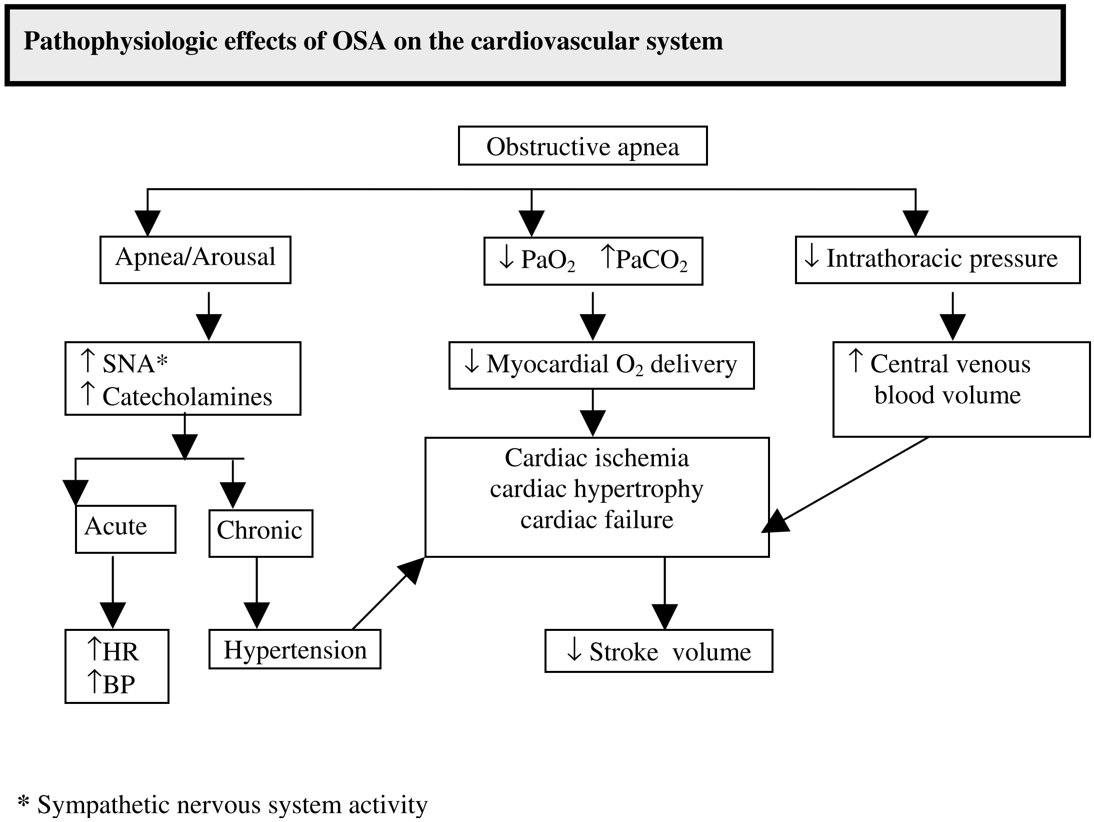


Figure 6–2 *Pathophysiological effects of obstructive sleep apnea syndrome on the cardiovascular system. BP, blood pressure; HR, heart rate;* $PaCO_2$*, arterial carbon dioxide partial pressure exerted by carbon dioxide (*CO_2*) dissolved in arterial plasma and red blood cell water;* PaO_2*, arterial oxygen partial pressure exerted by oxygen (*O_2*) dissolved in arterial plasma and red blood cell water; SNA, sympathetic nervous (system) activity.*

decrease after apnea. Cardiac output declines during apnea and increases after the resumption of ventilation. Intrathoracic pressure as low as −80 cm of water occurs during obstructive apnea as a result of the inspiratory efforts. This increases the central venous blood volume during apnea, with resulting increases in both central venous pressure and intracranial pressure. The augmented venous return to the right heart induces a leftward shift of the ventricular septum that in turn reduces left ventricular filling. The low intrathoracic pressure can also cause esophageal reflux.

SLEEP ARCHITECTURE

There are two kinds of sleep: non–rapid eye movement (N-REM) and rapid eye movement (REM). Within N-REM sleep, three stages are recognized, stage I, stage II, and delta sleep, corresponding to the different depths of sleep. REM sleep, of which there are also subcategories, is considered an important part of the overall sleep architecture where there is active dreaming during sleep. REM sleep alternates with N-REM sleep throughout the sleeping period.

A normal person will usually start the resting period with N-REM sleep stage I, which is a transitional phase between full wakefulness and sleep, that usually lasts between 1 and 7 minutes and that is characterized by a decrease in reactivity to outside stimuli. Stage II follows and is marked by the appearance of electroencephalogram (EEG) sleep spindles and by K complexes. In this stage, mental activity consists of short, mundane, and fragmented thoughts; it usually lasts between 35 and 45 minutes, at which time the person will enter delta sleep, which is a deeper level of sleep. Here the characteristic EEG delta waves appear (2 Hz and higher amplitude). The amount of time spent in this stage varies with the age of the person and ranges from a few minutes to 1 hour; it then yields to stage II sleep.

About 70 to 90 minutes after sleep onset the first REM period of the night occurs. It usually lasts about 5 minutes and is by far the least intense REM period of the night. The second sleep cycle begins while stage II sleep redevelops after the first REM period. On some occasions, delta sleep reappears, but there is generally less delta sleep in the second cycle than the first. Following this, the second REM period of the night occurs about 3 hours after falling asleep and lasts for about 10 minutes.

Following the second REM period and until awakening in the morning, stage II sleep and REM sleep alternate in

90-minute cycles. In these latter sleep cycles, delta sleep is rarely seen, whereas REM periods become more intense and are longer toward the morning. The mean length of a REM period is ~15 minutes, but some may last for 1 hour.

Although the separation of sleep into mutually exclusive stages is convenient for understanding, sleep stages actually merge. Delta waves, for example, gradually become more abundant and gain amplitude following the onset of sleep. There is no clear threshold of delta sleep, except by arbitrary definition. Similarly, indices of REM sleep are strongest in the middle of the REM period, whereas the transition point between stage II sleep and REM is often difficult to define.

Of the N-REM stages, delta sleep is the deepest and stage I lightest (if it is sleep at all). However, REM sleep is not easily classified on a sleep-depth scale. By measuring the noise necessary to awaken a person from REM sleep, it appears that REM in humans is about as deep as stage II N-REM sleep.

Several physiological systems are profoundly influenced by the patient's awake, N-REM, or REM state. In some persons, these systems may work normally during wakefulness but abnormally during sleep. In patients with OSAS, respiration may be normal during wakefulness, but relaxation of the musculature in the upper airway during sleep may have the potential for lethal complications.

CLINICAL SIGNS AND SYMPTOMS OF OSAS

There are two cardinal manifestations of OSAS in adults: one indicates a disturbance during sleep, the other the resulting disturbance during wakefulness. The first is loud snoring that typically has caused the bed partner to move out of the bedroom and occasionally out of the house. This is unlike the quiet, steady snoring that usually does not interfere with family life. The snoring of obstructive sleep apnea is frequently a crescendo variety, indicating increasingly severe narrowing of the airway. It can also be just a series of snorts interspersed with ominous silence. The daytime symptom is severe sleepiness. The patient shows evidence of falling asleep in all permissive situations. Occasionally, sufferers deny that they are sleepy and say that they fall asleep only when they sit down or are "bored." The truth is that boredom does not cause sleepiness. Unfortunately, one of the permissive situations is driving, and so patients with OSAS are at high risk for accidents. These two symptoms of loud, intermittent snoring and daytime sleepiness are so important that any person having them should be considered to have obstructive sleep apnea syndrome until proven otherwise. Conversely, it is extremely unlikely that clinically important OSAS exists in a patient who has no problem with daytime alertness and whose sleep at night is completely quiet.

TABLE 6–1 SIGNS AND SYMPTOMS OF OBSTRUCTIVE SLEEP APNEA SYNDROME

Symptom	Frequency
Loud snoring	Almost always
Hypersomnolence	In most patients
Abnormal motor activity during sleep	
Obesity	Frequent, but not necessary
Personality changes, depression	
Impaired intellectual performance	
Hypertension	40% of patients
Nocturnal cardiac arrythmias	Frequent
Cor pulmonale (advanced cases)	
Morning headaches	
Nocturia	Frequent
Sexual impotence	

In taking a history from the patient, one must rely heavily on the bed partner's observation, in terms of both nighttime and daytime behavior. Patients clearly are not aware of snoring because it occurs during sleep. They may also be unaware of how sleepy they are. Sleepiness frequently comes on very gradually, and these patients forget what it is like to be fully alert. Patients should be asked more than just "are you sleepy?" They should be specifically asked about falling asleep under permissive situations such as reading, watching television, and driving. The other clinical features of OSAS are not seen in every patient and are listed in **Table 6–1**.

PHYSICAL EXAMINATION

In the examination of the adult patient, one looks specifically for any abnormalities involving the upper airway such as nasal obstruction, hypertrophied tonsils and adenoids, retrognathia or micrognathia, and tumors. The majority of patients with OSAS are obese. One must look for frequently associated diseases such as hypothyroidism, acromegaly, and amyloidosis. Hypertension is more frequent in patients with OSAS than in control populations. Most patients with OSAS have normal pulmonary function, but any abnormalities in the respiratory system increase the risk of severe hypoxemia during sleep. A small subset of patients with OSAS develop daytime hypoventilation, so one should be alert for any evidence of hypercapnia or right heart failure. A summary of things to look for is enumerated on the physical examination sheet (**Fig. 6–3**). In our

Weight	______________________	
Height	______________________	
BMI	______________________	
Neck Size	15, 15.5, 16, 16.5, 17, 17.5, 18, 18.5, 19	
Nose	______	Deviated nasal septum Right Left %_______
	______	Turbinate hypertrophy
	______	Other ______________________
Nasopharyx	______	Adenoidal tissue
	______	Constrictor muscles prominent
	______	Mueller (0, I, II, III)
Oropharynx	______	Uvula ______ Normal ______ Elongated ______ Very elongated
	______	Retrodisplaced soft palate
	______	Soft palate defect from previous UPPP, LAUP
	______	Prominent anterior and posterior pillars
	______	Tonsils obstructing the oropharynx grade I, II, III, IV
	______	Surgical absence of tonsils
	______	Redundant wrinkled posterior pharyngeal wall
	______	Prominent hypertrophic constrictor muscles
	______	Macroglossia (true or relative)
	______	Müller (0, I, II, III)
	______	Other ______________________
Hypopharynx	______	Vallecula obstructed, LTH (grade 0, I, II)
	______	Epiglottis retrodisplaced
	______	Other ______________________
Skeletal	______	Craniofacial anomalies
	______	Hypoplastic mandible
	______	Malocclusion (retrognathia)
Diagnosis	______	Habitual snoring ______ Rule out sleep apnea
	______	Upper airway resistance syndrome
	______	Sleep apnea (obstructive, central, mixed)

Figure 6–3 *Snoring and sleep apnea physical examination form. UPPP, uvulopalatopharyngoplasty; LAUP, laser-assisted uvulopalatoplasty.*

Table 6–2 Polysomnogram definitions

Classification	Quantification or Severity
Apnea	Cessation of breathing >10 seconds
Hypopnea	Reduction in airflow, fall in O_2 saturation, terminated by arousal
Hypoxemia	A greater than 2% decrease in oxyhemoglobin saturation associated with an apnea
Apnea/hypopnea index (AHI)	The number of apneas + hypopneas per hour of sleep
Respiratory disturbance index (RDI)	Similar to AHI
Diagnosis of Sleep Apnea	
Central	Cessation of airflow secondary to a lack of respiratory effort; the diaphragm does not attempt to move
Obstructive	Cessation of airflow in the presence of continued thoracic breathing movements and exaggerated inspiratory efforts
Mild	AHI of 10 to 25 with hypoxemia and/or arousals and/or cardiac arrhythmias
Moderate	AHI of 25 to 40 with hypoxemia and/or arousals and/or cardiac arrythmias
Severe	AHI of >40 with hypoxemia and/or arousals and/or cardiac arrythmias
Mixed	Start as central apneas and are followed by an obstructive phase

experience, it is important to keep in mind the classification of obstructions as described by Fujita et al (1981), in which they note obstructions in the oropharynx, hypopharynx, or both.

We have also noted a higher increase in hypopharyngeal obstruction, more specifically lingual tonsillar hypertrophy in OSAS patients who have had a previous tonsillectomy. Lastly, the degree and site of collapse of the airway can be determined using the Müller maneuver, in which the patient is asked to breathe against closed nose and mouth while observed with a fiberscope.

LABORATORY EVALUATION AND DIFFERENTIAL DIAGNOSIS

Polysomnogram

The cornerstone in confirming our diagnostic impression of sleep disordered breathing, which has snoring as one of its components, is a polysomnograph. This can be performed in a hospital laboratory as a polysomnogram (PSG), with multiple leads analyzing a great amount of parameters, usually 12 or more, including EEG, electrocardiogram (ECG), and electromyogram (EMG), or in an ambulatory setting, in which portable units measuring a limited number of parameters are secured to the patient in the comfort of his or her home. Either way, the importance is to be able to define various components of sleep, in particular periods of apnea, hypoxemia, and cardiac arrhythmias, that when placed in a formula are able to confirm or exclude the diagnosis of habitual snoring, upper airway resistance syndrome (UARS), OSAS, and central sleep apnea. Polysomnogram definitions are listed in **Table 6–2**. The differential diagnoses of other sleep disorders are listed in **Table 6–3** and at times may require other specific tests for definitive diagnosis. The scope of this chapter will remain on obstructive sleep disorder breathing syndromes.

Snap Test

This inexpensive and simple test has been available since 1998 and is still in the process of acquiring clinical acceptance. The test is basically an acoustic analysis of snoring that can be performed singularly or in combination with an oxygen saturation monitor. The proponents of this

Table 6–3 Disorders of excessive somnolence

Psychophysiological
Psychiatric disorders
Use of drugs and alcohol
Sleep-induced respiratory impairment
- Sleep apnea syndrome
- Alveolar hypoventilation syndrome

Nocturnal myoclonus and restless legs
Narcolepsy
Medical, toxic, and environmental conditions
Other
- Kleine-Levin syndrome
- Menstrual associated syndrome
- Insufficient sleep
- Sleep drunkenness

diagnostic test claim that they can with a significant degree of accurateness differentiate those patients with habitual snoring from those with sleep apnea syndrome. Furthermore, this simple recording of snoring with a nasal cannula can quantify the loudness of snoring (in decibels) and localize the anatomical site or sites of snoring (by analyzing the pitch), determining if the sound originates from the nose, soft palate, tongue base, hypopharynx, or chest. We have found this test to be accurate when combined with clinical findings.

The Snap test (from the SPAP® laboratories) is one of several commercially available systems designed to be used in the patient's home to analyze the degree and character of sleep apnea. These systems have greatly improved in recent years, and a test with one of them can often replace a conventional PSG.

Radiological Examination

Cefalometry is the most widely available and inexpensive method for evaluating the skeletal and soft tissues of the head and neck. This two-dimensional modality provides useful information in patients with skeletal deformities, such as retrognathia, and is of aid in evaluating the usefulness of dental and lingual appliances. Computed tomographic (CT) scan examination, particularly spiral CT, provides direct three-dimensional volumetric reconstruction images of the airway and bony structures. It is useful in evaluating the efficacy of dental appliances and maxillomandibular advancement in patients with sleep apnea. At the present time the gold standard of imaging studies is dynamic magnetic resonance imaging (MRI). Although expensive, it has distinct advantages over other imaging modalities. Dynamic MRI provides the clinician with excellent airway and soft tissue resolution and an accurate determination of the upper airway cross-sectional area and volume. Because it is void of the risks of radiation, the studies could be repeated during wakefulness and sleep. Three-dimensional reconstruction of soft tissue structures (tongue, soft palate, fat pads, lateral pharyngeal walls) and the airway is possible.

CLASSIFICATION OF OSAS

The importance of the diagnostic tests is their ability to classify and quantify the severity of sleep disorders (**Fig. 6–1**). The classification of the degree of apnea is important because it correlates with the overall success of treatment. Patients with snoring, UARS, and mild OSAS have an overall better prognosis than those patients with moderate to severe OSAS. The various modalities of therapy have a predictable reduction in the apnea/hypopnea index (AHI) and will be discussed in the appropriate section.

The diagnosis of a specific sleep disordered breathing disorder is derived by a careful history and office questionnaire plus the physical findings and diagnostic testing. A brief discussion of each entity will follow.

Habitual Snoring

Identified as a partial obstruction of the upper aerodigestive tract without having a total collapse of the airway, characteristic habitual snoring is like a rhythmic seesaw whose loudness varies with individuals. Contributing factors to the frequency and loudness of snoring include excessive tiredness, heavy meals prior to retiring for sleep, and the use of sedative pills or alcohol. The patient may or may not have an increase in BMI. The patient does not report symptoms of daytime somnolence, and bed partners do not report patients gasping for breath. Polysomnography will determine the patient to have an AHI index of less than 10.

Upper Airway Resistance Syndrome

Patients with UARS present with symptoms of snoring that are accentuated by contributing factors including weight increase and alcohol and sedative drug use. The snoring is characterized by a crescendo pattern with associated arousals. After arousals there is a decrease in the upper airway resistance and temporary abolition of snoring. The multiple arousals lead to fragmentation of sleep with some tiredness throughout the day.

Polysomnography will demonstrate an RDI between 5 and 15, with no or occasional mild oxygen desaturations. Discussion exists whether this is a real entity or an intermediate stage in the spectrum of patients with sleep apnea. The only way to really identify an increase in the airway pressure is with an extensive 12-lead PSG, which will include esophageal pressure monitoring as one of its recording parameters.

Sleep Apnea Syndromes

Patients with sleep apnea syndromes usually have an increase in BMI of more than 20%. Hypertension or a family history of hypertension is present in more than half of the patients with this diagnosis. Easy fatigability and daytime hypersomnolence are characteristic symptoms, particularly in the patients with moderate to severe apnea. In the case of obstructive and mixed sleep apnea, the patient or a family member has witnessed a pattern of sleep with severe snoring, with cessation of breath for long periods of time and gasping for air upon arousal from sleep. Conversely, central sleep apnea is characterized by the absence of respiratory efforts during the periods of apnea. The differentiation between central, obstructive, and mixed sleep apnea is done by polysomnographic studies.

TREATMENT OF OSAS

The treatment of obstructive sleep apnea syndrome and sleep-related disorders is multifaceted and requires a careful analysis of the individual factors contributing to this entity. In some patients hypertrophied tonsils may be the single contributing factor causing the syndrome, whereas in the majority there are several factors that have been discussed previously. The point is that all contributing factors must be addressed for the best possible outcome. The severity of sleep apnea may itself be a predictor of outcome. Statistically, patients with mild to moderate sleep apnea will respond better to treatment than those with severe sleep apnea, regardless of the number of therapeutic modalities employed.

Medical Treatment

Weight control, continuous positive airway pressure (CPAP), and oral appliances are the major vectors in medical management of OSAS. Other medical conditions such as acromegaly and hypothyroidism, which can contribute to OSAS, have to be ruled out. Medications or substances such as alcohol, sedative hypnotics, narcotics, anesthetics, and sedating antihistamines, all of which have a depressive effect on the central nervous system, should be avoided. Control of hypertension is important in reducing the risk of cardiovascular complications in patients suffering from OSAS.

Weight Control

Obesity, most specifically the presence of a fat neck, is a major risk factor for the development of obstructive sleep apnea. Weight loss, therefore, is the cornerstone of treatment in every overweight patient, even those only mildly overweight. This treatment can be curative by itself, even with a minimal amount of weight loss. Studies have shown that as little as 10 to 15% reductions in weight can be associated with a 50% reduction in the number of apneas and a clinically significant improvement.

Continuous Positive Airway Pressure

In the majority of cases initial treatment of OSAS should be performed by CPAP. The equipment used consists of a mask or cannula that attaches to the nose, mouth, or both, and a generator that provides the delivery of air under pressure, which in turn acts as a "pneumatic stent" to prevent upper airway collapse. The amount of pressure required to maintain an open-air passageway varies depending on the severity of the disease and the collapsibility of the airway. Each patient's pressure level must be determined individually. The standard procedure is to observe the patient while sleeping and titrate the pressure to a level that eliminates apnea and snoring in all body positions and sleep stages.

There are conflicting reports regarding the long-term patient compliance with this form of treatment. The best long-term reports have a compliance rate of 70%; however, only one third of the patients adhere to the prescribed regimen when scrutinized. The reasons for abandonment of CPAP are varied, such as interference with the patient's lifestyle (i.e., intimacy, travel) and as a result of complications developing from the treatment itself (i.e., nasal congestion with rhinorrhea, facial irritation from the mask, eustachian tube dysfunction, and aerophagia with gastric distention).

Mandibular Advancement Devices

A mandibular advancement device (MAD) is a promising new approach in the treatment of snoring and obstructive sleep apnea syndrome. Satisfactory results are found more frequently in patients with mild OSAS than in those with a severe syndrome. The dental appliance pulls the lower jaw and thereby also the base of the tongue forward, which increases the size of the upper airway in the hypopharynx. The appliance also causes an increased tension in the oropharyngeal muscles.

Because the device applies strong force to the teeth and the temporomandibular joint, it is important that this type of treatment be performed by a dentist, preferably a specialist in orthodontics. Each device has to be made individually, and possible side effects must be carefully followed. Devices that are occluded, and hence prevent the possibility of oral breathing when the patient opens the mouth, are not recommended. Before intervention with a MAD, a clinical examination of the stomatognathic system including the mandibular joint has to be performed. To use an oral appliance, the patient has to have sufficient teeth and cannot have severe periodontal disease or cariogenic problems.

TABLE 6–4 SURGICAL PROCEDURES FOR SNORING AND OSAS

Nose and nasopharynx
Septoplasty
Turbinectomy
Adenoidectomy
Oral cavity and oropharynx
Tongue suspension suture
Tonsillectomy
UPPP
LAUP
Hypopharynx
Midline lingual glossectomy
Linguoplasty
Excision of lingual tonsil
Hyoid suspension
Partial epiglottectomy
Skeletal
Genioplasty
Genioglossus muscle advancement
Maxillomandibular advancement
Mandibular osteotomy
Other
Radiofrequency wave ablations
Tracheotomy

LAUP, laser-assisted uvulopalatoplasty; OSAS, obstructive sleep apnea syndrome; UPPP, uvulopalatopharyngoplasty

Other mechanical treatments include the use of nasal dilator splints and nasopharyngeal stenting, which bypasses the nose and palate.

SURGICAL TREATMENT

The numerous surgical procedures available for the treatment of snoring and OSAS (**Table 6–4**) underscore the difficulty in selecting the proper one. It is important to have a systematic approach when selecting the surgical procedures to be employed. Identifying the site or sites of soft tissue obstruction, skeletal deficiencies, and the severity of the disease in the spectrum of sleep-related disorders will help in selecting the most appropriate surgical procedures to be employed. An assessment of what is to be accomplished by the surgical procedure and the order in which it is approached should be thought out prior to embarking on this mode of therapy. Nasal and oropharyngeal surgery should be considered in the majority of cases because it usually correlates with the assessment prior to surgery. Skeletal surgery that mobilizes bone and soft tissue attached to it should be used as a primary modality only in cases in which gross craniofacial and skeletal deficiencies exist. Otherwise it should be performed as a staged procedure after other primary soft tissue surgery has failed and in patients with AHI indexes of greater than 25, as proposed by Powell et al (1990).

The aim of the surgical techniques is to enlarge and stabilize the airway. The mechanisms in which these techniques attempt to reach their objectives are by reducing or displacing soft tissue volume. Skeletal surgery, particularly the maxillomandibular advancement, will have the largest impact on the RDI scores. Future surgical techniques may focus on improving the neuromuscular tone of the tongue and pharynx by attempting to pace the neuromuscular units with external pacers. Early work on this frontier has been performed by attempting to pace the hypoglossal nerve. This experimental work may be expanded in attempting to pace the dilator muscles of the pharynx.

Some frequent surgical techniques employed in the treatment of sleep disordered breathing, and some salient points will be outlined below. The reader is encouraged to consult other textbooks and articles that specifically describe the techniques in detail.

Tracheostomy

This is the most effective surgical technique that bypasses the upper aerodigestive tract. It is 100% effective in relieving obstruction and correcting oxygen desaturation, hemodynamic alterations, and symptoms. However, there is a low patient acceptance of this procedure because it requires chronic maintenance and carries with it long-term morbidity at the tracheal site. A tracheostomy is indicated in patients with severe sleep apnea that is not respondent to CPAP and in patients with the alveolar hypoventilation syndrome.

Uvulopalatopharyngoplasty

Uvulopalatopharyngoplasty (UPPP) is a procedure in which the uvula, part of the soft palate, and redundant pharyngeal tissue are removed. The procedure increases the pharyngeal air space and stiffens the pharyngeal wall, decreasing the collapsibility of the upper airway. UPPP is effective in eliminating snoring in ~80% of patients, presumably by removing the tissues that vibrate. However, the surgery is not as effective in curing obstructive sleep apnea, achieving an overall success rate of ~50%. More dramatic responses are noted in those patients in whom hypertrophied tonsils are removed at the same time. Patients with a UARS and mild apnea have a better and more sustained response rate than those with moderate and severe sleep apnea syndrome, particularly if the patient is able to maintain or reduce his or her premorbid weight.

Serious complications from UPPP are infrequent, and postoperative bleeding will occur in less than 1% of the patients. Most of the complications are in relation to an overzealous attempt at removing excessive amounts of palatal tissue. Nasopharyngeal regurgitation of liquids will occur as a temporary symptom in a significant number of patients, particularly if drinking fast or leaning forward in the water fountain position. This symptom usually dissipates in a period of 3 months. Another sequela that the patient must be informed about is the sensation of oropharyngeal dryness, which in our experience occurs in one third of patients on a temporary basis, but in a few it will be present for extended periods of time. A rather frequent complication after UPPP, and to a lesser extent also after laser-assisted uvulopalatoplasty (LAUP), is swallowing dysfunction characterized by a loss of control of the bolus during the pharyngeal phase of swallowing. It is crucial that airways are closed when food is passing the hypopharynx, otherwise aspiration can occur. Only a few patients will actually have problems with aspiration after UPPP, but as many as 20% will have swallowing problems in that they have to concentrate very hard on swallowing to avoid aspiration when eating or drinking, which makes it impossible for them to talk while eating. There is evidence that this swallowing dysfunction may exist in snorers already before surgery, probably because of destroyed sensory nerve endings, but it may become overt after surgery. An infrequent but annoying complication is pharyngeal stenosis, where the patient complains of dysphagia, an inability to clear nasal secretions, and impaired breathing. Correction of this problem may require staged procedures with the use of lasers, stents, and pharyngeal flaps.

Laser-Assisted Uvulopalatoplasty

LAUP, an office-based procedure, reduces soft palatal tissue, the amount determined according to the judgment of the individual surgeon. The technique employed by most surgeons removes a triangular section of tissue adjacent to each side of the root of the uvula, followed by a reduction of ~50% of the distal uvula, thus shortening and elevating the size and position of the uvulopalatal complex. Both authors usually perform a sweeping technique in which the entire uvula and a portion of the soft palate are removed, leaving a surgical defect that looks like an arch, much like the defect left after a standard UPPP. LAUP is an effective modality in treating snoring and the symptomatology of UARS and mild OSAS, but it has little effect in changing the AHI.

Radiofrequency Tissue Volume Reduction

The proponents of this technique claim that by inserting electrodes into various portions of the soft palate and applying thermal energy, the soft tissue will experience a definable “thermal lesion” that over a period of 6 weeks will contract and be replaced by fibrotic tissue that “stiffens and reduces” the targeted soft tissue. The vibratory capacity of the palatal tissue will diminish, and as a consequence snoring is reduced. This procedure can be repeated multiple times and in multiple target sites of the upper airway, including the tonsils and tongue base. To date few reports have documented the effectiveness of this experimental therapeutic modality.

Tongue Base and Hyoid Bone Suspension

This is a recent technique in which a biocompatible screw with two nylon sutures attached to it is drilled into the lingual cortex of the symphysis of the mandible. The nylon attachments are then submucosally sutured, wrapped around the tongue base, and tied to each other in the floor of the mouth with a certain degree of tension. The proponents of this technique claim that the tongue base is contained from “dropping back” and narrowing the posterior airway space (PAS) during sleep. A similar technique is performed in which the attachment of the biocompatible screw to the mandible is the same; however, the free nylon edges are tunneled subcutaneously to the anterior portion of the neck and wrapped around the hyoid bone and tied under tension. This causes anterior and superior displacement of the hyoid bone, resulting in enlargement of the PAS. These techniques are performed on their own or simultaneously, and they can also be combined with other known procedures in the treatment of OSAS. To date fewer than 100 cases have been reported. Because the procedure is usually combined with other known surgical procedures in the treatment of OSAS, it is difficult to assess the amount of change in snoring and RDI scores by this technique. Furthermore, these “ingenious” novel techniques have yet to stand the scrutiny of time.

Maxillofacial (Skeletal) Surgery

Included in the category of skeletal surgical procedures are the inferior sagittal myotomy with hyoid bone suspension and maxillomandibular advancements. Maxillofacial surgeries increase the size of the upper airway by moving the base of the tongue away from the posterior hypopharyngeal and oropharyngeal walls, decreasing the collapsibility of the airway. The sleep apnea center at Stanford University (Powell et al, 1990) is one of the few to report findings using this surgical approach. Patients there are selected on the basis of the severity of their apnea (moderate to severe), presence of craniofacial abnormalities, such as micrognathia or retrognathia, or failure to respond to other therapy or as staged treatment in a surgical algorithm that performs this kind of surgery after having had prior nasal and oropharyngeal surgery. In Powell et al's series, maxillomandibular advancements combined with sagittal mandibular advancements have the greatest affect on OSAS. Unfortunately, their good to excellent results have not been matched by other centers performing the same kind of surgery.

Complications include those seen in other surgeries such as pain, bleeding, and wound infection. Problems specific to these procedures include dental nerve anesthesia, mandibular stress fracture, and esophageal reflux from decreased upper esophageal sphincter tone.

Goals of Treatment

Improvement of symptoms of daytime somnolence, fatigue, snoring, reduction in AHI, and nocturnal hypoxia are the ultimate goals of treatment. The physician and patient must understand the limitations of the various surgical procedures to be selected. Procedures like palatal somnoplasty and LAUP are aimed at reducing the amount of snoring, whereas UPPP and skeletal procedures are aimed at reducing the apnea/hypopnea index. More often than not UPPP does not significantly reduce the AHI scores; however, a good percentage of patients, for reasons that are not clear, will report improvement in their energy level and decrease in symptoms of fatigue and daytime somnolence. This is also true in patients undergoing an LAUP after having undergone a previous tonsillectomy or when the tonsils do not appear to be obstructing the airway. Also, patients with UARS respond well to the modalities mentioned.

SUMMARY

Although OSAS is very common within the American and European populations, this syndrome as an entity has only been known for 4 decades, first described by Gastaut in 1966. Treatment with CPAP and the UPPP operation both started in 1981, before which the only possible treatment was tracheotomy. Today we know a lot more of the serious effects of this syndrome, and there are more treatment modalities available, but this is still a "new" disease with many questions to be answered. One important question is what mechanisms underlie the pattern of progression from snoring to OSAS.

When taking care of the individual patient, the complexity of the syndrome must be kept in mind. Each patient, and if possible also his or her bed partner, has to be thoroughly questioned about both night and day symptoms, and the physical examination has to include an ear, nose, and throat examination as well as an examination of the heart and lungs. Before treatment such as CPAP, surgery, or dental appliance is started, an overnight sleep recording that will give an AHI is mandatory. The different possibilities of treatment have to be discussed with the patient, who also should be given information of possible side effects. Whatever treatment modality is chosen, each patient also has to be informed of the necessity of not gaining weight. After treatment, the effect of the treatment has to be controlled regarding the patient's symptoms and the outcome measured by the change of AHI.

The doctor treating patients with OSAS should have good knowledge of the complexity of this syndrome as well as of the whole spectrum of treatment modalities that are available. It seems preferable if the surgeon, the physician, and the dentist work together to optimize the best treatment for each individual patient suffering from heavy snoring and OSAS.

Suggested Readings

American Electroencephalographic Society Guidelines: guideline for polygraphic assessment of sleep-related disorders (polysomnography). J Clin Neurophysiol 1994;11:116–124

American Sleep Disorder Association. Intrinsic sleep disorders. In: Thorpy MJ, ed. The International Classification of Sleep Disorders: Diagnostic and Coding Manual. Lawrence, KS: Allen Press; 1990:52–61

Bålfors EM, Franklin KA. Impairment of cerebral perfusion during obstructive sleep apneas. Am J Respir Crit Care Med 1994;150:1587–1591

Findley L, Unverzagt M, Guchu R, et al. Vigilance and automobile accidents in patients with sleep apnea or narcolepsy. Chest 1995;108:619–624

Franklin KA, Nilsson J, Sahlin C, Näslund U. Sleep apnea and nocturnal angina. Lancet 1995;345:1085–1087

Friberg D, Ansved T, Borg K, et al. Histological indications of progressive snorers disease in an upper airway muscle. Am J Respir Crit Care Med 1998;157:589–593

Fujita S, Conway W, Zorick F, et al. Surgical correction of anatomic abnormalities in obstructive sleep apnea syndrome: uvulopalatopharyngoplasty. Otolaryngol Head Neck Surg 1981;89:923–934

Gastaut H, Tassinari C, Duron B. Polygraphic study of the episodic diurnal and nocturnal (hypnic and respiratory) manifestations of the Pickwick syndrome. Brain Res 1966;2:167–186

Gislason T, Benediktsdottir B, Bjornsson JK, et al. Snoring, hypertension, and the sleep apnea syndrome: an epidemiologic survey of middle-aged women. Chest 1993;103:1147–1151

Guilleminault C, Tilkian A, Dement WC. The sleep apnea syndromes. Am Annu Rev Med 1976;27:465–484

Haraldsson P-O, Carenfeldt C, Lysdahl M, Tingvall C. Does uvulopalatopharyngoplasty inhibit automobile accidents? Laryngoscope 1995;105:657–661

Hung J, Whitford E, Parson R. Association of sleep apnea with myocardial infarction. Lancet 1990;336:261–264

Jung R, Kuhlo W. Neurophysiological studies of abnormal night sleep and the pickwickian syndrome. Prog Brain Res 1965;18:140–159

Keenan Bornstein S. Respiratory monitoring during sleep: polysomnography. In: Guilleminault C, ed. Sleeping and Waking Disorders: Indications and Techniques. Menlo Park, CA: Addison Wesley; 1982:183–212

Levring Jäghagen E, Berggren D, Isberg A. Swallowing dysfunction related to snoring: a videographic study. Acta Otolaryngol 2000;120:438–443

Marklund M, Franklin KA, Sahlin C, Lundgren R. The effect of a mandibular advancement device in patients with obstructive sleep apnea. Chest 1998;113:707–713

Newman J, Clerk A, Moore M, et al. Recognition and surgical management of the upper airway resistance syndrome. Laryngoscope 1996;106:1089–1093

Powell NB, Riley RW, Guilleminault C. Maxillofacial surgery for obstructive sleep apnea. In: Guilleminault C, Partinen M, eds. Obstructive Sleep Apnea Syndrome: Clinical Research and Treatment. New York: Raven Press; 1990: 153–182

Riley RW, Powell NB, Li KK, Troell RJ, Guilleminault C. Surgery and obstructive sleep apnea: long-term clinical outcome. Otolaryngol Head Neck Surg 2000;122:415–421

Sullivan CE, Issa FQ, Berthom-Jones M, et al. Reversal of sleep apnea by continuous positive airway pressure applied through the nares. Lancet 1981;1:862–865

Self-Test Questions

For each question select the correct answer for the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. Below are listed some medical and surgical treatment modalities. These treatment modalities are usually performed for other reasons than affecting obstructive sleep apnea syndrome (OSAS). Four of them probably will have a beneficial effect on a person with OSAS, whereas one probably will worsen OSAS. Which treatment modality probably will worsen OSAS?
   A. Gastric banding
   B. Treatment of hypothyreosis
   C. Nasal corticosteroids.
   D. Pharyngeal flap as part of cleft palate repair
   E. Evulsion of nasal polyps
2. Which of the following statements is false?
   A. OSAS has no effect on intellectual performance.
   B. OSAS is more common in males than females.
   C. OSAS is associated with increased incidence of myocardial infarction and stroke.
   D. Diminished pharyngeal muscle tone will aggravate symptoms of snoring and sleep apnea.
   E. The most frequently occurring daytime symptom of OSAS is sleepiness.
3. Which of the following statements is true?
   A. Laser-assisted uvulopalatoplasty is recommended as treatment for a person with OSAS and an apnea/hypopnea index (AHI) of 46.
   B. Hypertrophic tonsils occasionally may be the single contributing factor causing OSAS.
   C. Obesity is not a major risk factor for development of OSAS.
   D. Uvulopalatopharynoplasty is more effective in the treatment of sleep apnea than in the treatment of snoring.
   E. Continuous positive airway pressure is not recommended for patients with a body mass index higher than normal.

Chapter 7

MICROBIOLOGY, VIROLOGY, AND MECHANISMS OF INFECTION

Ruy Soeiro and Bettie Steinberg

The respiratory system represents the largest surface area of the human body in contact with the external environment. This epithelial cell–lined structure is directly exposed to the environment, and thus in contact with a myriad of microbes as part of normal respiration. The nose, nasopharynx, and mouth normally are colonized with bacteria. However, the sinus structures, middle ear, and lungs are normally sterile. Because inhalation results in the deposition of microorganisms on the surface of the respiratory tract, a complex system of defense mechanisms has been defined that explains, to a great extent, the sterility of these sites. These defense mechanisms protect the human host. In this chapter, we will focus on the bacteria that cause infection of the upper respiratory tree—the posterior nasal cavity, nasopharynx, oropharynx, and sinuses—and, to a lesser extent, the middle ear.

SYSTEMS OF PROTECTION AGAINST BACTERIA

There are two systems that control bacterial colonization and infection of the upper respiratory tract. One is the innate defense system, the other the acquired immune defense system. The innate defenses are nonspecific mechanisms that are constitutively present. The posterior two thirds of the nasal cavity, the nasopharynx, and the sinuses are primarily lined by ciliated, pseudostratified epithelial cells. The epithelial linings also contain mucus-secreting goblet cells, found at high density in the nasal passages and to a lesser extent in the sinuses. Seromucous glandular cells are abundant in the nose but relatively rare in the sinuses. Mucus and fluid produced by the goblet cells and seromucous glandular cells, together with secretions from the submucosal vascular bed, form a blanket that is transported by the concerted action of the ciliated epithelium from the nasal passages and sinus cavities to the oropharynx. Microorganisms trapped in the mucous layer are thus delivered to the oropharynx and swallowed. The mucous layer, which travels at ~4.6 to 12.3 mm per minute, and which is replaced by newly produced mucus 2 or 3 times each hour, efficiently cleanses these cavities mechanically and helps to ensure sterility of the nasal sinuses. Other constituents of the nasal secretions, including plasma proteins in the mucin layer, inflammatory mediators, and antibodies or cells of the immune system, may also play a role in maintaining nasal sinus sterility. The volume and composition of the human nasal secretions thus binds microbes mechanically, exerts antimicrobial action, and prevents binding of microbes to cells of the upper respiratory tract. These functions are similar to those that protect the lower respiratory tree from invading pathogens. All act in concert as a primary host defense to protect the mucosal lining.

If the innate defense mechanism is compromised, infection can occur. Acquired mucociliary abnormalities have been shown in children with acute viral respiratory infections such as influenza and adenovirus type 1, and many bacterial infections of the respiratory tract and middle ear follow viral infection. Primary acute alcohol ingestion and smoke inhalation, ciliary dyskinesia, Young's syndrome, and cystic fibrosis all cause alteration of mucociliary transport that appear to play a role in decreasing host resistance to bacterial infection of the upper airway.

Up to 58% of patients with human immunodeficiency virus (HIV) infection experience recurrent sinus infections. Recent studies of these patients have shown impaired mucociliary transport due to increased viscosity of the mucus. However, defects in the innate defense system are not the only reason for increased infections in this population. Early in the course of HIV-induced disease, impaired resistance may reflect altered immunoglobulin synthesis as part of polyclonal B cell activation. Late in infection, severe immunoincompetence of both immunoglobulin and cell-mediated immune defenses are more common. The most common pathogens of immunocompetent hosts are also the most common in HIV-associated sinus infections. The less common gram-negative bacterial infections seen in these patients, including *Pseudomonas aeruginosa* and *Mycobacterium kansasii*, as well as opportunistic fungi such as *Aspergillus fumigatus* and *Candida albicans*, probably reflect a combination of altered clearance and altered immunity, as well as frequent use of antibiotics.

Finally, nosocomial sinusitis linked to nasotracheal obstruction is found in critical care settings. These infections are often due to gram-negative bacteria, selected for by antibiotic therapy, depressed consciousness, and steroid therapy so common in acute critical care situations.

The acquired immune defenses do not act constitutively. Rather, this system involves the antigen-specific responses of immunoglobulins and cell-mediated immunity that are induced in response to invading microbes. As a part of the innate defense system, and in response to signals of inflammation generated by the lymphoid system, granulocytes are mobilized and chemotactically attracted to the upper respiratory tract to act as phagocytes against many bacterial and/or fungal pathogens. By active phagocytic action, the cells ingest and, after phagolysosomal fusion, kill microbes. This is accomplished by a series of intracellular myeloperoxidases and amines that form powerful antimicrobial compounds that function together with granulocyte defensins, proteases, and hydrolases to destroy microorganisms that penetrate the mucosal barrier.

The chemoattraction of these cells, plus the increase in secretion of antimicrobial factors, and finally the stimulation of immunocytes to respond with antibodies, complement, and so on, in upper respiratory tract secretions, may be amplified and orchestrated by cytokines. These substances, produced by many cell types, play an active role in the secretion of innate factors of inflammation such as tumor necrosis factor α (TNF-α), the interleukins (ILs) 8,10, and 12, interferon γ, and granulocyte-macrophage colony–stimulating factor (GM-CSF). The details of the interaction of these molecules, the mechanisms of their function, and their potential role in the pathogenesis of viral and bacterial infections of the upper respiratory tract are beyond the scope of this chapter. However, cytokines appear to play an important role as an intrinsic part of both viral and secondary bacterial pathogenesis of sinus and middle ear infections.

SPECIFIC BACTERIAL INFECTIONS

The nonsterile areas of the upper respiratory tract are colonized by bacteria that are the indigenous flora of the adult nose. *Staphylococcus aureus* is found in the nasal vestibule of 25 to 40% of adults. The posterior nasopharynx of adults harbors *Streptococcus pneumoniae* (15–25% of people), *Haemophilus influenzae* (6–40%), *Streptococcus pyogenes* (6%), and other *Staphylococcus* species (12%). These are also the major pathogens found in infected sinuses of adults. *Moraxella (Branhamella) catarrhalis* is a frequent etiologic agent of acute infections in children, and gram-negative bacterial infections are often seen in adults. Many infections are negative for bacteria when cultured by routine methods. Most of these are undoubtedly due to virus infections, but *Chlamydia pneumoniae, Mycoplasma pneumoniae,* and *Legionella* are suspect organisms as well. Fungi are occasional causes of upper respiratory infections, especially in cases of diabetes mellitus or immunosuppression by either disease or medical intervention. In the next sections we will discuss some of these organisms in more detail.

Streptococcus Pneumoniae

This gram-positive aerobe is of overwhelming importance in infections of the upper respiratory tract. For this reason, it has been studied intensively. In fact, the study of the pneumococci has played an important role in the development of the field of molecular biology, as well as elucidating much of our understanding of the virulence of microbes. Studies of pneumococcus provided the first clear evidence that deoxyribonucleic acid (DNA) was the genetic material. Experiments showed that DNA was able to transfer the information required for capsular synthesis from encapsulated to noncapsulated strains. Research into understanding of serological types has clarified much of the epidemiology of this pathogen. Work on the mechanism of antibiotics elucidated the molecular target of penicillin, as well as concepts of tolerance to cell wall antibiotics. More recently, studies of the virulence of pneumococcus have led to new concepts regarding the role of the external capsule, the binding of the organism to human cells, and the role of the bacterial cell wall as a factor in human disease. The knowledge gained from study of this bacterium overshadows our understanding of other human pathogens and has pointed the way toward a greater comprehension of bacterial virulence in general.

The process of colonization and pathogenesis will be described in some detail for the pneumococci. Specific information on other common organisms that cause respiratory tract infections, as well as ways they differ from the pneumococci, will then be discussed.

Colonization

We now know that, for pneumococcus, as well as for many other microbes, binding to a human cell is the initial requirement for disease. Studies of the pneumococci have revealed that they may exist in two major phase variation types as distinguished by colony appearance: transparent and opaque. These variants possess different affinities for nasopharyngeal epithelial cells. The transparent colony types bind to host-cell sugar residues with between 10^2- and 10^6-fold more efficiency than the opaque types. There is strong evidence that glycoconjugate molecules on the membrane of the host cell serve as receptors because disaccharides can compete experimentally for binding to the cell surface. The primary site of binding of pneumococcus is the nasopharynx. In both children and adults, colonization may last up to several months without symptoms. It has been estimated that up to 90% of children may serve as carriers in the first 2 years of life.

Once bound, the colonizing organism may multiply locally to establish a stable colony. In the case of a capsulated strain, the organism is able to avoid ingestion by polymorphonuclear (PMN) cells that are attracted toward it for elimination. The capsule, although necessary for bacterial survival at this phase, does little to incite an inflammatory response by the host.

Pneumococci can also penetrate deeper into host tissues. Whether such penetration occurs by passage through the epithelial cells by endocytosis or between cells is unknown, but either mechanism may be operative as shown by studies of other pathogens. Antibody response to bacterial capsular polysaccharide occurs with colonization or asymptomatic invasion and is protective, maintaining colonization without evidence of disease. Exposure to a new serotype, or inflammation caused by a viral upper respiratory infection, may alter the host–parasite relationship and lead to acute infection.

Inflammation and Pathogenesis

In vitro, activation of epithelial cells by cytokines IL-1 or TNF-α (two of the cytokines produced during inflammation) alters host-cell surface molecules. These cytokines increase the density of platelet activating factor (PAF), which facilitates the binding of the bacterium, thus furthering colonization and increasing the likelihood for penetrating to deeper host tissues. This series of events is especially true for pneumococcal pneumonia, and presumed true for systemic bacteremia as well.

The cell wall, a macromolecule composed of several glycopeptides, has potent biological inflammatory activity. Isolated and purified pneumococcal cell wall is capable of eliciting an inflammatory response in the lung, the cerebrospinal fluid, and the middle ear. By analogy, initiation of inflammation of the paranasal sinus by pneumococcal cell wall constituents would also appear likely.

Initially, cell wall–induced inflammation causes separation of epithelial and vascular endothelial cell–cell connections. This results in vascular engorgement, leakage of serous fluid, attraction of PMNs, and induction of the inflammatory cytokines TNF-α and IL-1, all manifestations of acute bacterial-induced inflammation. It also activates complement by both classic and alternate pathways. Bacterial viability is unnecessary to induce these events. Killed encapsulated strains are capable of eliciting the full inflammatory response. Pneumolysin, a 53 kD bacterial toxin, is released as the host inflammatory response lyses the bacteria, resulting in further activation of inflammation, further toxicity to host cells, and amplification of the effects of the bacterial infection. Indeed, the concept that modulating the host response to pneumococcus-induced inflammation with corticosteriods, to minimize the effects of the infection, has now been incorporated into clinical treatment of pneumococcal meningitis in children.

In summary, the pneumococcus, once having achieved a successful colonization, is capable of being trapped in the closed spaces of the nasopharyngeal sinuses or the middle ear. This could be caused by inflammatory swelling of the osteomeatal complex or the eustachian tube, perhaps following a viral infection. Once there, release of cell wall substances results in additional inflammation, further perpetuating the damage done by host response mechanisms. The organisms may be localized by the host to the tissues involved initially, or may extend to the lung or the blood stream in a potentially fatal infection.

Haemophilus Influenzae

First identified as a pathogen by Koch, this facultative gram-negative, often pleomorphic, bacillus has long been identified as an important human pathogen. Generally described as an encapsulated bacterium, serotyping has identified six types of *H. influenzae* based on capsule antigenicity. The most important type clinically is type B. It causes most invasive diseases, including meningitis, pneumonia, cellulitis, sepsis, arthritis, and osteomyelitis. However, nonencapsulated strains are of great importance in infections of the ear, nose, and throat, particularly otitis media, sinusitis, epiglottitis, and conjunctivitis. Nonencapsulated strains also can cause chronic bronchitis and pneumonia.

Primarily found as a colonizing agent of the human respiratory tract, the nonencapsulated organism is carried asymptomatically for months at a time in the throats of 60 to 90% of infants and young children. In contrast, carriage of the encapsulated type B strain is less than 5%. Rates of colonization are lower for older children and adults.

Children under age 6 months have passively acquired maternal antibodies to type B, and only after age 2 does a child's immune system produce protective antibodies against the *H. influenzae* capsule. The period between ages 6 months and 2 years constitutes the time of greatest incidence of type B–induced disease. Recently, vaccination with capsular type B polysaccharide conjugated to a protein carrier has successfully immunized children in this age group. This has resulted in a major reduction of serious *Haemophilus* type B–induced disease as well as colonization in children.

Microbial transmission occurs from an infected (colonized) patient to the nasopharyngeal mucosa of a nonimmune individual. Several modes of binding have been identified. Specific sugars can mediate binding of *H. influenzae* fimbriae (specific bacterial cell surface organelles) to the mucous layer. However, nonfimbriated bacteria can also bind to the mucous layer. *Haemophilus* also binds directly to the surface of epithelial cells by mechanisms that have not yet been determined. Both capsulated and noncapsulated strains bind efficiently.

Once binding has occurred, the capsule appears to be of major importance in maintenance of colonization. *H. influenzae* is protected from secreted antibodies by bacterial IgA-specific proteases, and the capsule provides resistance to host polymorphonuclear leukocytes. The capsule is also important for invasion. Experimental studies of translocation across respiratory epithelium show that organisms whose capsule contains polyribosylribotol phosphate have pathogenic properties lacking in those with capsules of other serotypes or noncapsulated strains. It is this process of binding, followed by local invasion of the nasopharynx, that results in infection of the uvula and epiglottis. If the invading *Haemophilus* reaches the bloodstream, capsule-mediated resistance to phagocytosis allows bloodstream survival, high-grade bacteremia, and invasion of other tissues.

As with other microbes that predominantly attack by the respiratory route, effective mucociliary clearance is of major protection. Therefore, diseases that affect cilia, as well as environmental influences such as passive smoking, reduce this protection. As with the pneumococci, mucosal edema and impaired mucociliary clearance caused by viral infections result in trapping of *Haemophilus* in sinus and middle ear cavities, in turn resulting in bacterial sinusitis and otitis media.

Like other gram-negative organisms, all strains of *Haemophilus* express lipopolysaccharide (endotoxin). Along with other cell wall components, endotoxin induces an inflammatory response that both impairs ciliary function and disrupts the respiratory epithelium. It is in these areas of damaged epithelium that invasion appears to occur, resulting in both local inflammation and bacterial seeding of other host tissues.

In summary, *Haemophilus* organisms, both capsulated and noncapsulated, are able to bind to upper respiratory tract mucus and epithelium. If mucociliary clearance is compromised, or if the epithelium is damaged, the bacteria survive and successfully colonize the mucosa, further preventing successful mucociliary clearance by phagocytic host defenses. Invasion of the respiratory epithelium may be facilitated by a host inflammatory response to virus infection or by immune deficiency, as with acquired immunodeficiency syndrome (AIDS). Transport to lung tissue, especially in patients with chronic obstructive pulmonary disease (COPD), allows access to anatomically altered pulmonary tissue. This successful human pathogen thus has evolved to invade sterile tissue, by overcoming both static and acquired host defenses, to result in several well-recognized pathological consequences in both children and adults.

Moraxella (Branhamella) Catarrhalis

This gram-negative coccus resembles *Neisseria* species and was classified as *Neisseria catarrhalis* for many years. However, DNA studies showed a lack of genetic relatedness, and it has now been classified as a separate genus (*Moraxella*) within the family Neisseriaceae. The genus is now split into two subgenuses: *Branhamella* (coccal organisms) and *Moraxella* (more rod shaped).

These organisms are so common in the human oral flora that they have been considered normal commensals. However, they clearly can be etiologic agents of sinusitis, otitis media, mastoiditis, and pneumonia. Colonization of the nasopharynx occurs in up to 7.4% of adults and is increased in winter months. In contrast, 50% of children between 3 and 12 years of age carry *Moraxella*. This value is slightly higher (63%) in children with otitis media.

Transmission occurs from patients with upper respiratory infections to colonize the susceptible (nonimmune) individual. An increase in susceptibility to *M. catarrhalis* infection is apparent in disorders of humoral immunity. However, most adults have both immunoglobulin (Ig) G and IgA *Moraxella*-reactive antibodies that correlate with protective immunity. A rise in both serum and local antibody levels to bacterial outer membrane proteins (OMPs) is detectable with *M. catarrhalis*–documented otitis media. However, complete protection may not occur in otitis-prone children because colonization is so frequent in this group.

Moraxella, like other gram-negative organisms, possess lipopolysaccharide (endotoxin), which induces inflammation. Localized diseases, such as otitis media and sinusitis, are common with this bacterium, but disseminated disease is rare. This implies either that humoral immune resistance is sufficient to curtail most invasive infections or that the reduced frequency of invasion is due to lack of the virulence factors seen in *Pneumococcus* and *Haemophilus*. However, tracheobronchitis and pneumonia do occur, especially in older adults with COPD. This would suggest that *Moraxella* does have at least some potential for virulence.

Other Bacteria

S. pneumonia, H. influenzae, and *M. catarrhalis* are the most frequent community-acquired infections causing bacterial sinusitis and otitis media. However, there are several other bacteria that are human pathogens of the upper respiratory tract. These less often encountered organisms are usually found in chronic upper respiratory tract infections, infections of immunocompromised hosts, and nosocomial infections. They usually are found as polymicrobial infections, not necessarily preceded by viral upper respiratory infections (URI), and complicated by prior use of multiple antibiotics, immunocompromising drugs, and/or mechanical obstruction by nasotracheal or nasogastric intubation. In all of these circumstances, comorbidity of disease and other medications is common, and the ability to respond with innate as well as acquired host defenses is diminished. We will briefly discuss a few of these organisms in the next section.

Staphylococcus Aureus

This aerobic gram-positive coccus normally colonizes the mucosa of the anterior nares, but it can spread from there past the mucous membranes to contaminate other areas of the body. It binds to nasal mucus, facilitating its colonization and resulting in high rates of colonization (up to 30%), which may be either intermittent or prolonged. It does not bind to nasal mucosal cells, but rather to endothelial cells of the vascular system.

S. aureus is particularly common in patients with surgical or traumatic breaks in mucosa or skin, or decreases in polymorphonuclear leukocytes. These infections are often local, with a tendency toward abscess formation, but can become systemic with invasive and "metastatic" potential to bone, muscle, and heart. The organism possesses a capsule and secretes a variety of enzymes (coagulase, hyaluronidase, protease, nucleases) and toxins (hemolysins, toxic shock, toxin, and epidermonecrolysin) that are associated with systemic syndromes. Added to these virulence factors, the recent development of several forms of antibiotic resistance, along with the occurrence of *S. aureus* in compromised hosts due to intravascular and prosthetic devices, results in both severe local and systemic disease. Not clearly understood is its ability to establish chronic infections. This is presumably the result of multiple depressed local host defenses as well as its ability to persist in host phagocytes.

Acute ear, nose, and throat infections are rarely seen. Chronic infections, especially of nasal sinuses and middle and external ear, are more common. Infections are frequently mixed with anaerobic and gram-negative organisms that are also part of the normal flora of the mouth. However, *S. aureus* is usually the sole pathogen in acute parotid gland infection, especially in the elderly.

Pseudomonas Aeruginosa

This aerobic gram-negative bacillus is not normally a part of the endogenous flora of the human host. It has gained notoriety as an opportunistic pathogen, adhering through pili to gangliosides on human mucosa and other cell surfaces. A ubiquitous organism associated with warm, moist environments, it also possesses multiple virulence characteristics that facilitate human infection. Endotoxin, common to all gram-negative bacterial cell walls, can induce destructive host inflammation. In addition to endotoxin, it possesses a series of exotoxins. Exotoxin A affects protein synthesis of human cells through adenosine diphosphate (ADP)-ribosylation of elongation factor 2, in precisely the manner found for *Corynebacterium diphtheriae*. This toxin causes local cell death. In synergy with its co-cytotoxin, a pore-forming toxin of human cells, exotoxin A impairs host PMN function. These toxins, and the proteases and hemolysins secreted by this organism, facilitate invasion.

Commonly associated with plastic catheters, and selected for by surgical tissue breakage, *P. aeruginosa* is a frequent nosocomial organism. It causes local infections of paranasal sinuses of noncompromised hosts and severe infections in patients compromised by the obstruction of nasotracheal tubes, impaired mucous membrane clearance, and neutropenia. These infections have become more difficult to treat because of the selection for antibiotic resistance in hospitalized patients. Classic infections of the ears, especially malignant external otitis, are seen in elderly diabetic patients, again in those with decreased normal host defenses. Simple external otitis, auricular perichondritis, and chronic infections of nasal sinuses or otitis media are also well described.

Klebsiella Species

Other facultative gram-negative bacilli share virulence factors of capsules, endotoxins, exotoxins, and invasiveness. Most *Klebsiella* infections are due to *K. pneumoniae*; however, of particular note is *K. rhinoscleromatis*, of which little is known about specific virulence factors. This organism is capable of chronic infections of the pharyngeal mucosa, perhaps associated with its large antiphagocytic capsule and its intrinsic resistance to intracellular killing. Often seen in immunocompromised hosts, it causes an indolent, chronic, granulomatous process that appears in upper respiratory tissue as a scarring, obstructing lesion, often resistant to prolonged antibiotics, and often requiring surgical removal. *K. ozaenae* causes atrophic rhinitis, a disease seen more in underdeveloped countries, particularly of Southeast Asia.

Anaerobes of the Oropharynx

Most bacteria in the normal flora of the human oropharynx are anaerobes, numbering between 10- and 100-fold more abundant than aerobes. These gram-positive and -negative organisms are associated with a variety of chronic infections of the ear and sinuses and of dental infections. They also cause complications by spreading to the lung and the brain. These infections are usually polymicrobial, also including many of the aerobes already described.

The organisms express several virulence factors that lend to their extension from the mouth. Many possess antiphagocytic capsules, endotoxins, toxins that attack host defenses, and exoenzymes capable of tissue destruction. The list of organisms, present in all standard treatises on oral microbes, includes *Peptostreptococci, Fusobacteria*, and *Prevotella* (previously called *Bacteroides melaninogenesis*). This collection is variously aerotolerant, often requiring the reduced pH and oxygen (O_2) tension characteristic of tissue necrosis, decreased blood supply, and obstructed cavities to grow. The combination of anaerobes and facultative aerobes may give rise to virulence synergy too complicated to dissect, but resulting in further tissue-destructive relationships.

These combinations of organisms are isolated from multiple infections of tonsils, sinuses, middle ear, and deep neck, as well as from oropharyngeal Vincent's angina and necrotizing ulcerative gingivitis. They may induce inflammation resulting in a host component to the pathology; may directly destroy tissues; facilitate spread of aerobes to the neck, brain, or lung; and prevent antibiotic effects because of the high frequency of β-lactamase production by this group. Together, they represent a significant pathogenic potential of the normal flora of the oropharynx, resulting frequently in postsurgical infected wounds, infected wounds of malignancies of the head and neck, and infectious complications due to mechanical trauma to the oronasopharynx. Infections are usually polymicrobial and require antibiotics against both aerobic and anaerobic species for their cure.

HOST RESISTANCE TO VIRAL INFECTIONS

The same innate and acquired immune defense systems that protect against bacterial infections also protect against viruses. Successful infection requires specific attachment of proteins on the viral surface to receptors on the target cell, followed by penetration into the cell. The receptor for a given virus can be limited to a specific cell type or tissue in one species of animal, providing highly restricted host specificity, or can be broadly represented on many cell types. The presence of a mucous barrier in the respiratory tract limits and reduces the likelihood of successful viral infection by blocking access to the receptors.

The interferon system can also be considered a part of the innate system of protection because it is not specific for antigens of a given virus. Synthesis of double-stranded RNA (dsRNA), usually late in the viral replication cycle, induces synthesis of type I interferons (interferon α from infected lymphocytes and interferon β from fibroblasts). Binding of type I interferons to uninfected cells then induces an antiviral state that limits successful spread of the viral infection (**Fig. 7–1**). Destruction of messenger RNA (mRNA) and phosphorylation of the translation factor F-2 inhibit both viral and host protein synthesis in the cell. In some cases, the inhibition simply reduces the ability of newly infected cells to produce more virus; in other cases, it kills the newly infected cells, thereby halting spread of infection. Patients who are unable to generate interferons suffer from severe disseminated infections. They can be treated by injections of purified interferon. However, the side effects of interferon (fever, malaise, hair loss) have limited its use as a therapeutic agent to severe infections such as hepatitis B and C.

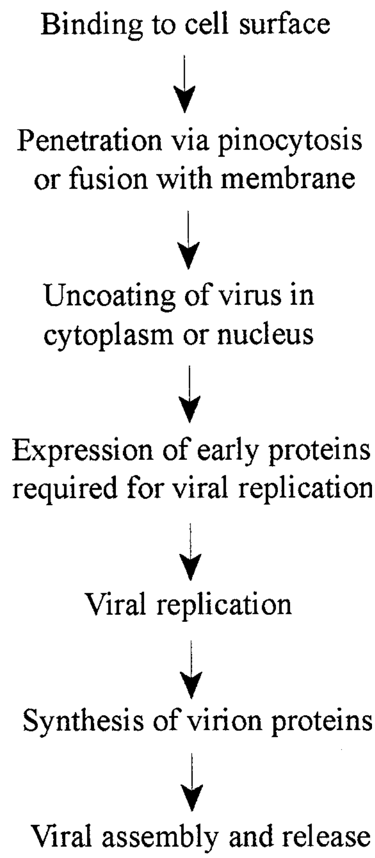


Figure 7–1 *Induction of interferon and establishment of an antiviral state in uninfected cells.*

Those viruses that kill cells slowly and generate large amounts of dsRNA during infection are strong inducers of interferon, whereas viruses that rapidly kill the host cell, prevent synthesis of host proteins, or do not generate dsRNA as part of their life cycle are generally poor inducers. Moreover, not all viruses are susceptible to inhibition by interferon. Adenovirus, one of the viruses associated with respiratory infections, is resistant to interferon, while other respiratory viruses such as rhinovirus and influenza virus are susceptible.

Antigen-specific acquired immune defenses, described earlier in this chapter, are key to preventing and managing most viral infections. Specific antibodies reduce viral load and infectivity, and antibody-based immunization has successfully limited or eliminated many viral scourges including smallpox, polio, and measles. However, antibodies are not essential to manage viral infections. Children with agammaglobulinemia primarily

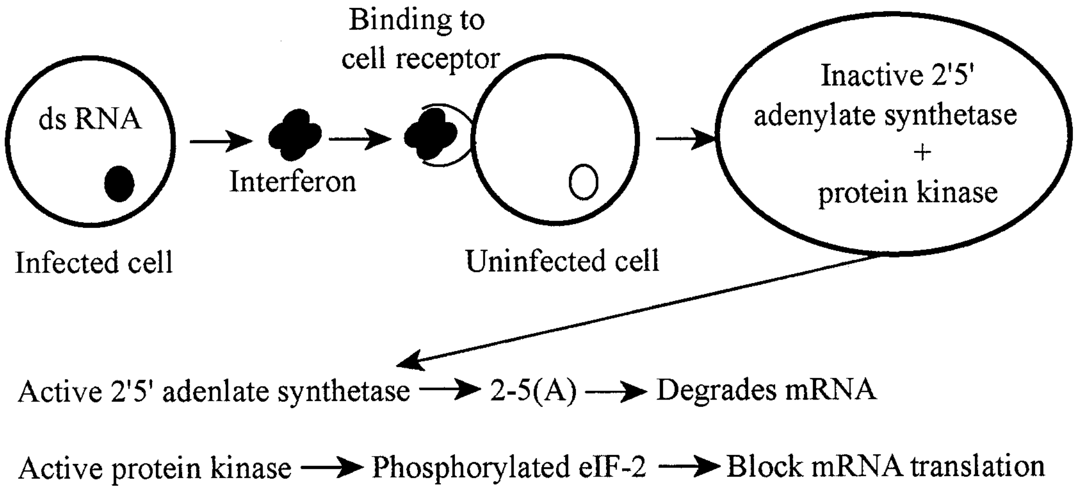


Figure 7–2 *Steps in the life cycle of an active virus infection.*

suffer from bacterial rather than viral infections. Antigen-specific cell-mediated immunity, on the other hand, appears essential to control viral infections. Patients with DiGeorge's syndrome, characterized by congenital thymic aplasia, suffer from severe and frequent viral infections. Measles, for example, is often fatal in these patients. Some viruses have evolved ways to evade the immune system, permitting long-term persistent or latent infections. For example, adenoviruses and herpesviruses produce proteins that block presentation of major histocompatibility complex (MHC) class I molecules on the cell surface, preventing effective cytotoxic T-cell destruction of the infected cell.

OVERVIEW OF VIRAL STRUCTURE AND REPLICATION

Viruses have a much simpler structure than bacteria, consisting of nucleic acid, usually complexed to one or more proteins, enclosed within a protein coat or capsid to form a nucleocapsid. Subsets of viruses are surrounded by a lipid envelope derived from cellular membranes. The structure of the nucleic acid varies among viruses and is a major factor in determining classification. The genome can be either DNA or RNA, linear or circular, in one piece or several, double or single stranded. The capsid protects the nucleic acid from the environment and mediates entry into the host cell. It is normally composed of several subunits or capsomeres, each made of one or more proteins that are assembled to provide a specific shape to the viral particle. Most capsids are either icosahedral (20 sided) or helical. A few viruses (e.g., poxviruses) have complex multilayered protein coats surrounding the nucleic acid, rather than a simple capsid. The lipid envelope is derived from cellular membranes. The envelope is studded with virally coded glycoproteins, some of which interact with cell-surface receptors, and matrix proteins that link the envelope to the nucleocapsid. Presence of a lipid envelope increases the sensitivity of the virus to solvents such as ether.

Virus infection is a multistep process (**Fig. 7–2**). A successful infection requires that the virus enters a host cell and subverts that cell to producing viral proteins and nucleic acids, assembling new virus, and releasing that virus to repeat the cycle. An abortive infection, for example, would occur if the virus bound to a cell that could not support viral replication. There are several steps in the cycle where the process can be blocked. Antiviral approaches that inhibit binding (such as soluble receptors that "soak up" virus or small peptides that bind to cell-surface receptors and compete for virus binding) are attractive preventive strategies currently being developed. Once a virus has penetrated, antibodies and agents that inhibit binding are no longer effective. Replication is dependent on a combination of viral and host enzymes. Expression of viral proteins required for replication can be inhibited by interferon (see above). In the small DNA viruses, nearly all of the host DNA replication machinery is used, and a small number of virally encoded proteins provide enhanced viral DNA replication over cellular replication. In the poxviruses and herpesviruses, which are very large, the viruses code for many replication enzymes, and dependence on host enzymes is reduced. Retroviruses, such as HIV, code for an RNA-dependent DNA polymerase (reverse transcriptase) that copies the viral RNA into copy DNA (cDNA)—the opposite of the normal transcription, where DNA is transcribed into RNA. The viral cDNA is inserted into the host cell chromosome and then transcribed by host RNA polymerases to make many copies of viral RNA. Other RNA viruses either code for their own RNA polymerase or alter the cellular polymerase by addition of viral subunits. These are all targets for antiviral therapy.

The previous process describes the infectious cycle of viruses. Some viruses establish latent infections, where the virus binds, penetrates, is uncoated, and then fails to replicate for a protracted period of time. The herpesviruses are the best-studied examples of latency. Herpes simplex initially infects epithelial cells of the oral mucosa or genital tract and establishes a small localized active infection. The virus then travels along nerve fibers to the dorsal root ganglia, where it can persist for the life of the patient. In the ganglia, only a few latent-specific RNAs are expressed, and there is no viral replication. If local trauma or inflammation of the epithelium occurs, some of the viral DNA travels back down the nerve fiber to the epithelium, reenters the epithelial cells, and enters the full replication cycle.

Another variant of the life cycle is a persistent infection, with virus being produced at very low levels for a protracted period of time (often years). Many viruses can cause persistent infection, including measles, mumps, HIV, and Epstein-Barr virus. This type of infection is characteristic of the "slow viruses" that are associated with some neurologic diseases and is suspected of playing a role in multiple sclerosis.

PATHOGENESIS OF VIRAL INFECTIONS

Multiple factors influence the outcome of a viral infection. These include the patient's age and nutritional and immune status, as well as the specific type of virus. Some viruses cause asymptomatic or minor illness in young children but severe illness in adults. For example, poliovirus infection is usually asymptomatic in infants, especially in populations where the virus is endemic and most adults have antibodies that partially protect infants. The advent of better hygiene and delayed infection resulted in the polio epidemics of the 1940s and 1950s, which were especially severe in middle-class populations. Other viruses cause only minor illness in older children and adults but severe illness in infants (e.g., respiratory syncytial virus). The characteristic pathogenic processes of viruses commonly associated with diseases of the head and neck are discussed below with the individual viruses.

ANTIVIRAL THERAPIES

Historically, the best approach to control virus infections has been preventive—vaccines and better hygiene to provide safe food and water. More recently, several specific therapies have been developed to treat viral infections after they begin (**Table 7–1**). In addition, nonspecific treatments such as interferon are used for some types of severe infections. The basic philosophy behind most antiviral drugs is to inhibit an enzyme or process in the viral life cycle that is not shared by the host cell. For some viruses, this would appear relatively easy. Viruses that encode unique enzymes (e.g., RNA viruses that use reverse transcriptase to convert their nucleic acid to DNA) should be treated effectively by drugs that target the enzyme. However, most viruses use host enzymes for transcription and replication. Any drug that interferes with these processes will also interfere with normal host functions, causing significant side

TABLE 7–1 ANTIVIRAL DRUGS

Virus	Drug(s)	Mechanism(s) of Action
Influenza A	Amantadine	Blocks uncoating of virus, binds viral ion channel protein; only effective if given very early in infection
Herpesvirus	Acyclovir, gancyclovir,foscarnet	Phosphorylated by viral thymidine kinase; analogs incorporated into viral DNA block viral DNA polymerase; prevents viral replication in both initial and reactivated latent infections; binds to viral DNA polymerase, blocks activity
RSV	Ribavirin	Purine nucleoside analog blocks viral RNA polymerase
HIV	Azidothymidine, dideoxyinosine, dideoxycytosine,protease inhibitors	Nucleoside analogs block reverse transcriptase, act as chain terminators for viral cDNA synthesis, inhibit viral replication during viremia; if given early enough, can prevent initial establishment of infection; inhibit viral protease, required to process multigene transcripts into final RNA

cDNA, complementary DNA; DNA, deoxyribonucleic acid; HIV, human immunodeficiency virus; RNA, ribonucleic acid; RSV, respiratory syncytial virus.

effects. Moreover, the reverse transcriptase mentioned above is an error-prone enzyme. This means that the mutation rate for these viruses is very high, and treatment with antiviral drugs selects for those mutant viruses that are more resistant to the drug. This has been and continues to be a major problem for therapies for HIV infection. Despite these limitations, several effective antiviral therapies have now been developed.

RESPIRATORY VIRAL INFECTIONS

Several viruses cause respiratory infections in infants, children, and adults. Though usually self-limiting, these infections resulting in a large amount of absenteeism from school and work. Often, they are associated with secondary bacterial infections like those discussed earlier in this chapter. These infections can involve the sinuses, the middle ear, and the lower respiratory tract. Thus viral respiratory infections have a significant impact on the health care costs of the nation.

Rhinoviruses

Rhinoviruses are the cause of the common cold. They are small, nonenveloped single-stranded RNA viruses, closely related to poliovirus. There are over 100 different serotypes of rhinovirus, and antibodies against one type are usually not protective against another. The average adult will get one or two rhinovirus infections per year, and young children get six to eight. Rhinoviruses infect the ciliated epithelium of the nose, with virus replicating and shed for 3 to 5 days. The viruses are temperature-sensitive, unable to replicate at 37°C, which limits the infection to the nasal passages, which are slightly cooler than the rest of the respiratory tract. Rhinoviruses are highly infectious. Transmission usually occurs by transfer from contaminated hands touching mucous membranes of the eyes or nose. Symptoms are caused by the host inflammatory immune response, rather than by any significant damage to the mucosa itself.

Coronaviruses

Coronaviruses are a common cause of colds occurring in the late winter and spring, causing 10 to 25% of upper respiratory infections. They are also single-stranded RNA viruses, but unlike the rhinoviruses they are enveloped. These viruses are primarily associated with infections in adults rather than children. Transmission is normally by a fecal–oral route, rather than airborne droplets or from hands to nose. The pathophysiology of disease symptoms is very similar to rhinoviruses.

Influenza Viruses

Influenza viruses are quite different from either rhinoviruses or coronaviruses. The virus genome is composed of eight separate segments of RNA. Replication of this RNA virus, unlike many other RNA viruses, occurs in the nucleus. Replication is mediated by an RNA-dependent RNA polymerase coded for by the virus.

Influenza virus enters the nasopharynx, invades the mucous layer, and infects the ciliated respiratory epithelium. Infection causes necrosis and desquamation of the epithelium, resulting in a loss of the protective mucous layer. Many of the systemic symptoms of influenza are caused by release of cellular contents, including interferons, during necrosis. Because the protective ciliary action on the mucous layer is lost, the respiratory tract is highly vulnerable to secondary infections. Approximately 10% of acute influenza virus infections subsequently lead to pneumonia.

Immunization is a very effective preventive of infection, if the vaccine is against the same virus types and subtype currently in circulation. The influenza virus is surrounded by a lipid envelope with multiple protein spikes embedded in it. The virus is very unstable, with a high mutation rate. Mutation of the spike proteins causes a new subtype of virus, partially limiting effectiveness of antibodies from earlier influenza infection or immunization. Multiple infection of a single cell by two or more different types can generate a new mix of RNA segments. This results in a new viral type—the cause of worldwide epidemics with few people resistant due to prior infection with other types. Many of these new types originate in China. Both ducks and pigs harbor influenza viruses that can infect humans. Ducks and pigs are grown together in China, facilitating mixing of infections from one to another. One of the major problems in influenza vaccine development is predicting, a year in advance, which subtypes or new types that have appeared are likely to spread through the population.

Parainfluenza Viruses

Despite its name, this virus is not related to influenza virus. Parainfluenza viruses are single-stranded RNA viruses most closely related to the mumps virus. The virus replicates in the cytoplasm and buds through the cell membrane without killing the host cell. However, infected cells lose ciliated function, some cells are lost, and some fuse to form giant multinucleated cells. Parainfluenza virus types 1 and 2 cause laryngotracheobronchitis, and type 3 causes bronchiolitis and pneumonia. Transmission is by both aerosol and hand.

The infection is usually subclinical in adults but acute in children.

Respiratory Syncytial Virus

This single-stranded RNA virus is the most frequent cause of upper respiratory infection in babies. Approximately 1% of infants with clinical disease will require hospitalization. Respiratory syncytial virus (RSV) is transmitted by hand contact from adult carriers and is a common nosocomial infection. It has been estimated that 25 to 30% of the staff in a newborn nursery carry RSV. Adults with compromised immune systems are also at risk for RSV infection. Infection initially occurs in the upper respiratory tract, where it kills the infected epithelial cells, but in babies infection can spread to the lower tract, causing multiple symptoms, including croup, bronchitis, bronchiolitis, and pneumonia. T-cell response to infection is critical for control but can also contribute to the pathology, with a delayed hypersensitivity reaction in the lungs.

RSV is a rather fragile virus and not too easily infectious. The initial vaccine against RSV, tested several years ago, actually proved to exacerbate disease. Antibodies bound to the virus enhanced uptake by macrophages by binding to receptors on the cell surface. Recently, a new vaccine has been developed that should protect infants against RSV infection.

Adenovirus

The last of the viruses that commonly cause respiratory infections is the adenovirus. Unlike all those listed above, this virus contains DNA as its nucleic acid, with one moderately large double-stranded molecule containing 35,000 to 38,000 base pairs. (For comparison, papillomavirus DNA is only 8000 base pairs, whereas Epstein-Barr virus is 171,000 base pairs). The viral DNA replicates in the host cell nucleus, using host enzymes in combination with virally coded enzymes. Adenoviruses are nonenveloped, with fiber spikes projecting from the icosahedral capsid at each apex. These fibers determine the hemagglutination properties of the virus.

Adenoviruses are very stable, maintaining infectivity for several weeks at 4°C, and several months at −20°C. They are grouped into seven subgroups, subdivided into 40 serotypes. The pathogenesis of adenovirus infection varies, depending on the serotype. The most common otolaryngological disease manifestations are acute respiratory disease restricted to new military recruits (types 4 and 7), pharyngoconjunctival fever and pharyngitis (type 3), conjunctivitis (types 3 and 7), and epidemic keratoconjunctivitis (types 8 and 19). Several fatal cases of nonbacterial pneumonia in infants have been ascribed to types 3 and 7. Transmission is from person to person through respiratory and ocular secretions. Infection of epithelial cells causes a complete halt to cellular DNA and RNA synthesis, as the host machinery is subverted to synthesis of viral components, but the infected cells do not usually die. The viral material is only inefficiently assembled into new viral particles, and most accumulates in the nucleus as large inclusion bodies that identify the infected cells. Most adenovirus infections are subclinical, and most people are infected with one or more serotypes before the age of 15. Adenoviruses often establish latent infections in the tonsils. Infection with one serotype is generally protective against subsequent infection with the same serotype, and neutralizing antibodies persist for many years. The persistent elevated titers probably reflect occasional low-level activation of the latent tonsil infection, which serves to restimulate the immune system. Activation of latent infection can be a problem in immunosuppressed patients.

VIRUSES ASSOCIATED WITH OTOLOGIC DISORDERS

In addition to contributing to bacterial middle ear infections secondary to viral respiratory infections that cause swelling of the nasopharyngeal mucosa and obstruction of the eustachian tube, viruses are also directly associated with otologic disorders. These can result in permanent hearing loss.

Rubella Virus

This virus was associated with epidemics in the past that resulted in large numbers of congenitally deaf children. Fortunately, a vaccine against rubella has now been developed. Rubella is an enveloped, single-stranded RNA virus that is a member of the Togaviridae family. It is the only member of the family that exclusively infects vertebrates, while the other viruses in the family infect and replicate in insects for at least part of their life cycle. Rubella virus RNA replicates in the cytoplasm, viral proteins are inserted into the outer cell membrane, and the virus buds through this membrane for a prolonged period of time. Although the virus is cytoplasmic, cultured cells infected with rubella frequently show multiple chromosomal breaks and damage.

Infection normally occurs in the respiratory tract, transmitted by nasal secretions, but then causes a hematogenously spread viremia. The infection is generally self-limiting in children and adults, and normally causes little pathology in the host beyond a rash that may appear 14 to 25 days after infection. In fact, most young adults with rubella infection are asymptomatic. However, during the viremia that precedes the rash, rubella can cross the placenta and infect a developing fetus. If infection occurs during the first trimester of pregnancy, multiple congenital developmental abnormalities, including deafness, cataracts, cardiac abnormalities, and motor deficits, occur in ~30% of patients. Damage to the fetus is caused by rapid death of some fetal cells, coupled with persistent infection and chromosomal breaks and abnormal mitoses in other cells. Babies born with congenital rubella infection are infectious and can shed virus for up to 2 years.

Measles Virus

Measles virus can cause otitis media and mastoiditis. Measles is a member of the paramyxoviruses, related to parainfluenza virus, RSV, and mumps, with a single strand of RNA and a viral envelope. Like the other paramyxoviruses, the viral particle contains an RNA-dependent RNA polymerase that initiates viral replication. It is the most infectious known virus, with successful infection established by as little as one viral particle. The virus is transmitted by respiratory secretions, enters the upper respiratory tract, and replicates in respiratory epithelium and regional lymph nodes. This phase of infection can be characterized by conjunctivitis, dry cough, sore throat, headache, low-grade fever, and tiny red patches with central white specks on the buccal mucosa (Koplik's spots). The localized infection then spreads to more distant lymph nodes and skin by hematogenous dissemination, resulting in the characteristic rash. Multinucleated giant cells can be detected in nasal secretions and sputum from infected patients prior to rash appearance. The most frequent complications of measles infection are bronchopneumonia and otitis media. However, the most serious complications are neurological—encephalomyelitis and subacute sclerosing panencephalitis (SSPE). SSPE is a chronic, progressive fatal disorder due to persistent low-level infection.

Mumps Virus

The mumps virus, another paramyxovirus, can also cause otologic disease. The virus is transmitted by saliva and respiratory secretions, infects the respiratory tract, and replicates initially in respiratory epithelium and cervical lymph nodes. The virus then spreads to the salivary glands, the primary target organs, through a hematogenous viremia. This viremia occurs several days prior to clinical mumps symptoms, during which time the patient is infectious. Although disease is usually limited to the salivary glands, multiple other organs can be involved causing orchitis, meningitis, meningoencephalitis, and deafness. Mumps can cause deafness in the absence of other central nervous system symptoms, through destruction of the organ of Corti. Deafness is often preceded by a sense of "fullness" and tinnitus, and is bilateral in 20% of affected patients.

Human Immunodeficiency Virus

This virus, more commonly associated with its immunodestructive functions, can also cause otologic symptoms. HIV is a lentivirus, an enveloped retrovirus containing two identical copies of RNA. Other lentiviruses cause slowly developing diseases in sheep and horses. It has recently been nearly conclusively proven that HIV infections were transmitted from chimpanzees to humans, and probably occurred multiple times, generating different HIV strains in the human population. HIV infects T cells, macrophages, astrocytes, and neurons. The initial infection is latent, and activation requires both cellular and viral factors. HIV is characterized by a high rate of mutation, which has limited effectiveness of antiviral drugs and further complicated development of a successful vaccine.

HIV RNA has been found in outer hair cells and supporting cells of the cochlea and is associated with progressive high-frequency sensorineural hearing loss in 20 to 25% of AIDS patients. Active infection causes cell damage or loss through induction of apoptosis (programmed cell death), and adjacent uninfected cells can be killed by a bystander effect. Neuronal damage can be mediated through release of excitotoxins from infected macrophages and supporting cells. These toxins include nitrous oxide, arachidonic acid, and quinolinate. Moreover, there is some evidence that zidovudine (AZT), one of the primary drugs used to treat HIV infection, is itself ototoxic.

Cytomegalovirus

Cytomegalovirus (CMV) is a herpesvirus, one of the family that also includes herpes zoster and herpes simplex. Herpesviruses are very large, enveloped viruses that contain one linear double-stranded molecule of DNA. These viruses enter the cytoplasm, where they are uncoated, and the DNA then moves to the nucleus. Herpesviruses replicate and are assembled into new virus particles in the host cell nucleus, with replication dependent on a virally coded polymerase that is a target for antiviral therapy, and acquire their envelope as they bud through the nuclear membrane. The enveloped viruses then transit the cytoplasm and exit the cell.

Herpesviruses can cause both acute and latent infections. Infection with CMV is widespread, but disease associated with infection of adults or children is rare. However, congenital CMV infection, acquired in utero, has a different outcome. Twenty percent of infants with congenital CMV infection will display severe symptoms (primarily neurological), and 25% of these infections are fatal. Five to 20% of "nonsymptomatic" infants will have significant sensorineural hearing loss due to cochleitis and auditory neuritis.

TUMOR VIRUSES

A subset of head and neck tumors is caused by viruses, although the vast majority of tumors are caused by environmental factors, primarily tobacco and alcohol use. A viral etiology holds out the hope that appropriate antiviral therapies might eventually be useful for the treatment of some tumors. Unlike most of the viruses causing respiratory and otologic disease, which are RNA viruses, the two viruses that cause head and neck tumors are DNA viruses. They are Epstein-Barr virus (EBV) and papillomavirus. Both viruses cause acute and persisting latent infections, and only a rare subset of infections results in tumors. Tumorigenicity is multifactorial, involving the virus and other cofactors. These cofactors include environmental agents that can cause cellular mutations or damage and host factors such as immune response to the viral infection.

Epstein-Barr Virus

EBV is also a member of the herpesvirus family, with the same mode of replication and nuclear budding as CMV. Unlike the other herpesviruses, however, EBV establishes its latent infection in B cells and respiratory epithelium rather than in neurons. This latent infection is often located in tonsils, and tonsillar tissue will yield infectious EBV if cultured in the laboratory. The initial infection is usually asymptomatic, especially in young children. In teenagers and adults, initial infection is often manifested as infectious mononucleosis.

Nasopharyngeal carcinomas (NPCs) are caused by EBV. The virus infects the epithelium, establishes a latent infection, and expresses a subset of viral proteins including Epstein-Barr nuclear antigen 1 (EBNA-1) and EBNA-2. Detection of these proteins is diagnostic for latent EBV infection. Activation of the virus by cofactor(s) can then lead to induction of the cancer. The role of EBV in NPC was first recognized in China, where the southeast region (primarily Canton) has a very high rate of NPC. This geographic region coincides with the area where the population has a high rate of exposure to croton oil. We now know that both croton oil and tung oil can promote EBV-induced NPC. More recently, scientists have realized that NPC in other parts of the world, including the United States, is also caused by EBV. The low incidence of NPC in the West probably reflects the low exposure to cocarcinogens that specifically promote the tumors, rather than lower prevalence of latent EBV infection. The cocarcinogens responsible for NPC in the United States and Europe have not been defined.

Papillomaviruses

Human papillomaviruses (HPVs) are a family of ~100 closely related, very small nonenveloped DNA viruses that infect stratified squamous epithelium, including skin, respiratory, digestive, and genital mucosa. Viral type is based on DNA sequence homology, because there is no good serological test for HPV infection at this time. HPVs code for only eight to 10 proteins, including the two capsid proteins. The viral DNA replicates in the nucleus of infected cells, using the cell replication enzymes. Thus it is very difficult to prevent viral replication without significant toxicity to uninfected cells. The virus codes for one replication protein that enhances viral DNA replication, but to date there has been no successful antiviral therapy that targets this enzyme. Papillomaviruses are transmitted by contact with squamae shed from the surface of infectious lesions. The infection remains local, with no evidence for systemic spread.

The majority of infections are subclinical or latent. For example, ~5% of the population has latent respiratory tract infection, 30% of young adults have genital tract infection, and most people have latent skin infections. The most common manifestation of active HPV infection is a benign tumor, or papilloma. Eight to 10 HPV types are associated with papillomas in the oral cavity, oropharynx, nasopharynx, and larynx. Recurrent respiratory papillomatosis, caused by HPVs type 6 and 11, is the most serious benign disease caused by HPV infection. The disease is characterized by recurrent lesions that can be removed surgically but frequently recur. Recurrence is due to activation of latent infection, widespread in the respiratory tract of these patients, by some unknown process. Though unusual, the disease can involve not only the larynx but also the trachea, bronchi, and lung. Lung involvement is usually fatal. There is no effective medical treatment for respiratory papillomas and no cure. Interferon controls the disease in many patients, but the papillomas recur as soon as treatment stops. Most patients are unwilling to undergo years of interferon therapy because of its side effects.

Malignant conversion of benign HPV-induced lesions also requires cofactors. Radiation is a known cofactor, and is thus contraindicated in the treatment of respiratory papillomas. Other cofactors are unknown, but any agent that can cause DNA damage is likely to be able to induce malignant conversion. Because HPV infection (primarily with types 16 and 18) is the cause of more than 90% of cervical cancers, many investigators have sought a role for HPVs in head and neck cancers. These investigations have had contradictory results, most probably confounded by the presence of latent HPV infection in oral and respiratory mucosa. There is, however, mounting evidence that tonsillar cancer is caused by HPV, usually HPV 16. More than half of all tonsil cancers contain HPV 16, and the two viral oncogenes, E6 and E7, are expressed at high levels. These two viral genes are not expressed in latent infection. The restriction of HPV-induced head and neck cancer to tonsils raises the possibility that HPV and EBV might interact in the pathogenic process. This is currently under investigation by several research groups.

SUMMARY

Clearly, a large number of bacteria and viruses cause otolaryngological diseases and disorders. However, one must always remember that these disorders do not necessarily exist as monomicrobial infections. Humans are colonized with large numbers of bacteria, most of which do not cause disease even if pathogenic types, and carry many latent or persistent viruses. Mere presence of an organism does not necessarily define etiology of a disease. As our ability to detect smaller and smaller numbers of organisms increases, we risk overinterpreting results. Moreover, an active infection with one organism can directly contribute to a second pathogenic infection. This is most clearly seen in cases where an initial viral infection, for example, with influenza virus, leads to a secondary bacterial infection of the respiratory tract. It is the complex interactions of multiple factors, including the status of the host, that determine outcome of bacterial and viral infections.

Suggested Readings

Dulbecco R, Ginsberg HS, eds. Virology. 2nd ed. Philadelphia: JB. Lippincott; 1990

Fauci AS, Brounwald E, Isselbacher KJ, et al, eds. Harrison's Principles of Internal Medicine. 14th ed. New York: McGraw-Hill; 1998

Field BN, Knipe KM, Howley PM, eds. Fields Virology. 3rd ed. Vols 1 and 2. Philadelphia: Lippincott-Raven; 1996

Gwaltney JM. Acute community-acquired sinusitis. Clin Infect Dis 1996;23:1209–1225

Johnson JJ, Yu VL, eds. Infectious Diseases and Antimicrobial Therapy of the Ears, Nose and Throat. Philadelphia: WB Saunders; 1997

Joklik WK, Willett HP, Amos DB, Wilfert CM, eds. Zinsser Microbiology. Norwalk, CT: Appleton and Lange; 1992

Kaliner MA. Human nasal respiratory secretions and host defense. Am Rev Respir Dis 1991;144:S52–S56

Steinberg BM. The role of human papillomaviruses in benign and malignant lesions. In: Hong WK, Weber RS, eds. Head and Neck Cancer. Norwalk, CT: Kluwer Academic; 1995:1–16

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. Which one of the following answers is correct? Penicillin resistance is primarily due to
   A. Intrinsic high mutation rate of *Streptococcus pneumoniae*
   B. Selection of resistance in hospitalized adults
   C. Childhood pharyngeal colonization
   D. Frequent use of oral antibiotics for viral upper respiratory infections (URI) in children
   E. Potential for inadequate drug level in childhood otitis media

2. Pathogens causing acute community acquired bacterial sinusitis usually include all but one of the following:
   A. Alpha-hemolytic streptococci
   B. *Haemophilus influenzae*
   C. *Moraxella catarrhalis*
   D. *Staphylococcus epidermidis*
   E. Mixtures of anaerobic bacteria

3. The differential diagnosis of acute viral versus acute bacterial sinusitis is difficult because
   A. The conditions usually occur simultaneously.
   B. The signs and symptoms of the two conditions have considerable overlap.

C. In the absence of an air–fluid level, sinus x-rays or computed tomography (CT) examinations cannot reliable distinguish between the two conditions.
D. A therapeutic trial of antibiotics does not result in rapid clearing of sinus abnormalities on x-ray or CT examination.
E. All of the above

4. Viral respiratory infections are difficult to treat because
A. Most viral infections are not diagnosed early enough in the viral life cycle to initiate effective drug therapy.
B. Most viruses use host enzymes for replication, and thus there is a limited therapeutic index for antiviral drugs.
C. The high mutation rate of viruses causes them to be resistant to most antiviral drugs.
D. The overlapping symptoms of adenoviruses, rhinoviruses, and influenza virus infections preclude accurate diagnosis and thus choice of the proper antiviral drug.
E. Most people do not generate sufficient interferon titers.

5. Pathogenic outcome of viral infection is multifactorial, due to the following:
A. Host expression of interferon, causing systemic effects, is induced by only some viruses.
B. The age of the patient at the time of infection can alter symptoms.
C. Some viruses primarily cause latent or low-grade persistent infection.
D. Previous exposure to a virus can induce production of neutralizing antibodies that prevent or limit infection.
E. All of the above

Chapter 8
PRINCIPLES OF PHARMACOLOGY

CHRISTOPHER J. HARTNICK, ALEXANDER W. GOTTA, AND IRA M. LEVITON

This chapter is designed to present an overview of the pharmacokinetics necessary to understand the processes by which drugs of different varieties enter the human body, exert their effects, and are excreted. The subject of pregnancy as it relates to the pharmacology of drug administration will also be discussed. After this overview, several classes of drugs will be examined; namely, local anesthetics, antireflux medications, steroids, nonsteroidal anti-inflammatory drugs, analgesics, and antibiotics. The topic of antihistamines is covered in Chapter 3 on allergies.

PHARMACOKINETICS

Pharmacokinetics describes the process by which drugs are absorbed into the organism and eventually into the cell, distributed to the various compartments within the body, and ultimately excreted from the body. The activity of a given drug is dictated by the interplay of these processes.

The process of absorption actually begins not at the cell membrane but rather at the point of administration to the organism. Various factors affect this process: the solubility and concentration of the drug, the circulation at the site of absorption, the surface area for absorption and the route of ingestion. The critical step of the process of absorption is the drug's passage across the cell membrane. The cell membrane is a dynamic arrangement of a bilayer of lipids with protein molecules either embedded in one side or completely traversing the membrane. A drug's ability to pass through this membrane relates to the intrinsic properties of the drug as well as those of the cell membrane itself. The molecular size and shape of the drug, the solubility of the drug at the site of absorption, and the relative lipid solubility of the drug's ionized and nonionized forms are all important factors.

Drugs can cross the cell membrane by several routes; some require energy and are therefore active processes. Others do not require energy and are termed passive. Drugs that are electrically charged and have difficulty entering into a cell are transported by carrier-mediated active membrane transport. Their passage across the membrane is linked to the passage of certain anions and cations [usually sodium (Na) and potassium (K)] and requires adenosine triphosphate (ATP). A variant of this type of transport occurs when the charged particles do not have to travel against the established intercellular electrochemical gradient and therefore the process does not require energy. This process of carrier-mediated transport is called facilitated diffusion.

Drugs crossing cell membranes by passive mechanisms do so by passive diffusion along concentration gradients according to their solubility in the lipid bilayer. This diffusion is proportional to the magnitude of the concentration gradient across the membrane as well as to the lipid/water coefficient of the drug. It follows that the greater the coefficient, the higher the concentration of the drug in the membrane, and therefore the faster the diffusion.

Passive diffusion differs for drugs that are electrolytes from those that are nonelectrolytes. For nonelectrolytes, after steady state is achieved, the concentration of that drug on either side of the membrane is equal. For electrolytes, the concentration of that drug relates to the differences of pH across the membrane as well as to the pK_a. This is an important concept because most drugs are either weak acids or bases existing both in ionized and nonionized forms. *The Pharmacological Basis of Therapeutics* (Hardman et al, 2001) offers an excellent example of a drug that is a weak acid in the stomach where the gastric mucosa acts as a partition between the acidic environment in the stomach and the plasma itself. It is assumed for the purpose of this example that the mucosa acts as a lipid bilayer cell membrane.

In plasma, the ratio of the nonionized to ionized drug is 1:1000, whereas in the gastric contents the ratio is 1:0.001. An opposite result would be seen if the drug were a weak base. Understanding these concepts is extremely important because of the tremendous implications in terms of further absorption, and then upon excretion.

Distribution

After a given drug is absorbed into the body, it is distributed to the various compartments within the body. Depending on the solubility and concentration of the drug and on the drug-binding ability of plasma proteins, the drug can either go through the systemic circulation or be trapped by reservoirs such as fat and be released slowly over a period of time. The volume of distribution relates the initial dosage of a particular drug administered and absorbed to the serum concentration of the drug.

Biotransformation

Once the drug has been absorbed and distributed and has exerted its given effect, it must be altered in some fashion to allow the body a means of excreting it, or else it would continue to accumulate until reaching toxic levels. Moreover, some drugs reach their active state only after having been altered in the body. These two needs, of acilitating excretion and promoting an intended action of a given drug, are accomplished by the transformation of lipophilic drugs to more polar, less lipophilic substances. This is done by enzyme systems found mostly in the smooth endoplasmic reticulum of the liver (named the microsomal system), but also found in the kidney, lung, and gastrointestinal system. Drugs absorbed enterally or rectally are first presented for biotransformation to the liver; this is the so-called first-pass effect. Drugs that are administered sublingually travel directly to the systemic circulation from the superior vena cava and avoid the first-pass effect.

Biotransformation is a process of oxidation/reduction (phase I) reactions, conjugation reactions (phase II), or a combination of the two. The key step for oxidation reaction is the insertion of one atom of molecular oxygen into the substrate. This usually produces an unstable intermediary that will then be broken down further. This process of oxidation is largely controlled by a large family of isoenzymes called cytochrome P-450 that are located in the smooth endoplasmic reticulum of the liver. The most common factors affecting the biotransformation process by the P-450 system are genetically determined polymorphisms. A second factor affecting this process of biotransformation includes the administration of certain drugs such as ethanol and cimetidine that inhibit the P-450 enzymatic process. A third factor that can affect the cytochrome P-450 is hepatic blood flow.

Conjugation is the other means of biotransformation that facilitates excretion. Conjugation requires glucoronyl transferases that are found mainly in the hepatic endoplasmic reticulum but also can be found in the kidney and other tissues. Hyperbilirubinemia of infancy is due to a deficiency of this enzyme. Once conjugated, drugs are excreted into the bile and urine.

Excretion

Drugs are excreted mainly from the kidney, but also in the stool. In the case of anesthetic gases, excretion takes place via the pulmonary system. Importantly, drugs can be excreted into breast milk, and this must be remembered to prevent unwanted effects on the infant.

Renal excretion is a process of glomerular filtration, active tubular secretion, and passive tubular reabsorption. It is highly influenced by the pH of the urine. When the urine is made more alkaline, weak acids are excreted more rapidly; weak bases are preferentially excreted when the urine is made more acidic. Some of the protocols for the treatment of drug overdoses and poisoning derive from these tenets.

Many of the products derived from the hepatic first-pass effect are excreted into the feces; others are reabsorbed in the blood and then excreted in the urine. Alternative, though less consequential, routes of excretion include saliva, sweat, and tears.

By adding all the routes of excretion to achieve a single rate of excretion for the entire organism, one can begin to derive a dosing regimen that will not produce toxic levels. With the given assumption of complete bioavailabilty (a measure of how much of a dose gets into the body and reaches its site of action), at steady state, the rate of drug administration will equal the rate of drug elimination, and a dosing rate can be easily calculated. This concept of measuring the rate of elimination from the human body is called clearance. Clearance can be expressed as a measure of the total body, or it can be individualized for a particular organ system such as the kidney or the liver. Once the clearance and the volume of distribution are known, the half-life of a particular drug can be determined. It follows that as the clearance decreases, the half-life of a particular drug would increase and vice versa.

PHARMACOKINETICS AS IT RELATES TO PREGNANCY

Placental transfer of drugs depends on several factors: the physiochemical properties of the drug, the rate at which the drug crosses the placenta, the duration of exposure, the pattern of fetal distribution, and the stage of the placenta and fetus at the time of exposure. The general tenets of pharmacokinetics hold that the more lipophilic the drug, the more readily it will pass through the placenta. Of note, though, even polar substances will pass through the placenta given high enough concentration gradients.

Another factor influencing placental transfer is the molecular size of the drug. Drugs with a molecular weight of 250 to 500 readily pass through, whereas drugs with a molecular weight greater than 1000 have difficulty crossing. Protein binding also produces an effect. This is somewhat influenced by maternal blood flow and other factors, including lipid solubility and relative ionization.

Once entering the placenta from the maternal system, two mechanisms are present that help protect the fetus. First, the placenta functions not only as a semipermeable barrier but also as a site of biotransformation. Several different types of oxidation reactions have been demonstrated to occur in the placenta. Second, drugs enter the fetal circulation via the umbilical vein, where 40 to 60% of this blood enters the fetal liver and undergoes some degree of the first-pass effect. In addition, a large proportion of the blood from the umbilical artery is shunted back through the umbilical vein to the liver for another round of biotransformation. There are obvious limits to the amount of biotransformation possible in the fetal liver. For example, as mentioned earlier in the case of neonates and applying to the fetus as well, a deficiency of glucoronyl transferase in the fetal liver inhibits the process of conjugation. Nevertheless, the fetus is afforded a moderate degree of protection from these mechanisms of biotransformation.

A complete list of medications that can adversely affect the fetus is beyond the scope of this textbook. The reader is referred to the *Physicians' Desk Reference (PDR)* for a drug-by-drug accounting of the indications, contraindications, and dosages before prescribing a given drug to the pregnant patient.

SUMMARY OF PHARMACOKINETICS

Armed with the fundamental understanding of how drugs enter the human body, exert their effect, and then are executed, we are now ready to examine certain distinct classes of drugs used in the field of otolaryngology—head and neck surgery.

GASTROESOPHAGEAL REFLUX DISEASE: TREATMENT OF THE OTOLARYNGOLOGICAL MANIFESTATIONS

Approximately 10% of all laryngeal complaints are thought to be secondary to reflux esophagitis. The causative factors precipitating reflux esophagitis include an incompetent lower esophageal sphincter (LES), increased acid production or decreased clearance, alteration of the epithelial resistance to acid injury, and an incompetent upper esophageal sphincter (UES). The signs of laryngeal reflux esophagitis differ from the normal presentation because they are seldom accompanied by the classic esophageal symptoms of heartburn and regurgitation. The classic presentation of laryngeal reflux disease is a patient complaining of hoarseness, chronic cough, cervical dysphagia, or the sensation of a foreign body in the throat (globus sensation). Recent studies demonstrate a strong connection between laryngeal stenosis and reflux disease and hint at an association with laryngeal carcinoma.

The workup of suspected laryngeal reflux disease begins with a proper history and physical exam. The patient should be questioned as to the previously described symptoms. Initial examination should include indirect evaluation of the larynx: signs of edematous, erythematous arytenoids, or heaped up tissue in the posterior pharyngeal region should be considered highly suspicious for reflux disease. Double-probe pH monitoring is the gold standard diagnostic test, but for the case of suspicious reflux laryngitis in the absence of either stenosis or carcinoma, many treat empirically for the presumptive diagnosis.

The first-line treatment for laryngeal reflux laryngitis is avoidance of substances such as spicy foods, chocolate, tea, and caffeine, as well as smoking. The medical armamentarium to fight reflux disease includes proton pump inhibitors and H_2 blockers. Other agents that are used for classic reflux esophagitis that can be used as complementary agents include the classes of prokinetic drugs.

H_2 Receptor Blocking Agents

Cimetidine (Tagamet), ranitidine (Zantac), famotidine (Pepcid), and nizatidine (Axid) are all examples of the class of drugs that act as H_2 receptor antagonists. Acid production is stimulated by a combination of histamine, gastrin, and acetylcholine. The H_2 receptors appear to be specific for acid production in the human body. The receptor antagonists act by competitive antagonism and function by preventing acid and pepsid secretion. The major side effects of this class of drug are mostly associated with cimetidine and its interaction with the hepatic cytochrome P-450 system. Cimetidine has been shown to slow down the clearance of drugs using the P-450 system. The other H_2 receptor antagonists do not interfere with the P-450 system. Cimetidine also binds to testosterone receptors and thus can cause gynecomastia and decreased libido in males. Ranitidine, for its part, has been shown to reduce renal clearance of triamterene in humans. Overall, this class of drug is safe and is now available in over-the-counter preparations.

Proton Pump Inhibitors

Drugs such as omeprazole (Prilosec) and lansoprazole (Prevacid) function by preventing the parietal cell H^+, K^+-adenosinetriphosphatase (ATPase) from transporting H^+ into the lumen of the stomach. These drugs therefore prevent acid production at the final common pathway and are the most potent antacids available. The potentially serious side effect of significant gastric mucosal hyperplasia has failed to be demonstrated in humans.

Prokinetic Drugs

Drugs such as erythromycin (E-Mycin), cisapride (Propulsid), and metoclopramide (Reglan) act by increasing peristaltic contractions, by increasing LES resting tone, or by combining these two actions. Metoclopramide functions as a dopamine D_2 antagonist to increase the tone of the LES and to increase the force of and coordinate peristaltic contractions. It is also an antiemetic. Cisapride acts via the serotonin pathways as a 5-HT agonist, and also functions to increase the LES resting tone and to increase and coordinate contractions. Cisapride can cause diarrhea. It is metabolized by the P-450 system; therefore, its levels can be elevated when used with other drugs that saturate the P-450 system. A rare side effect due to cisapride overdosage is arrythmia.

Treatment Strategy

The largest prospective study concerning the treatment of laryngeal reflux disorders was performed between 1983 and 1988 by Kaufman (1991) and involved 225 consecutive patients who presented with a history either of laryngeal cancer (31 patients) or laryngeal stenosis (33 patients), or a history suspicious for gastroesophageal reflux disease (GERD, 143 patients); 18 patients were categorized as miscellaneous. Patients were initially treated with lifestyle alteration and medical therapy consisting of H_2 receptor antagonists administered for a duration of 6 months. After 6 months, 85% of the patients experienced relief of the symptoms with which they had initially presented. Over a 4-year period of time, 20% of the patients who had experienced relief after the initial treatment with medication had recurrences of the symptoms. The overall failure rate for medical treatment was 35%. Those patients who failed medical treatment went on to undergo Nissen fundoplication. After this landmark study, the advent of the proton pump inhibitors has changed the arsenal such that these agents are now the first-line medication for medical treatment of laryngeal reflux disorders. Large-series prospective studies with long-term statistics from this treatment strategy have yet to be published.

GLUCOCORTICOIDS AND THEIR USE IN OTOLARYNGOLOGY

Glucocorticoids are formed in the adrenal cortex, which uses cholesterol as a precursor substrate. Production of glucocorticoids is highly regulated by the hypothalamic-pituitary-adrenal (HPA) axis. Production is increased during times of physiological stress to the organism. Glucocorticoids exert their effects broadly; they promote gluconeogenesis in the liver and lipolysis in adipose tissue, and they affect protein synthesis and degradation in striated muscle. They regulate growth and development in fetal tissues and induce the production of surfactant in the fetal lungs. Their most important actions relative to the field of otolaryngology stem from their effect on the body's immunologic response to foreign insult either by injury or by inflammation from foreign antigen; it is this suppression of the immune system that renders glucocorticoids potent in the treatment of allergic, inflammatory, and autoimmune disorders. Glucocorticoids diminish the ability to adhere to vascular endothelium after injury. They inhibit the antigen processing by macrophages and suppress T-cell helper activity. They inhibit the cytokine response to injury or inflammation. They promote eosinopenia and lymphopenia. Glucocorticoids perform this immunosuppression at the price of decreasing the body's ability to ward off infection. This potentially disastrous side effect must be remembered each time glucocorticoids are to be prescribed for a length of time.

Glucocorticoids, as mentioned previously, are produced in the adrenal cortex as cortisol and then are released into the circulation. There, most of the cortisol is bound to cortisol-binding globulin, and 5 to 10% is bound to albumin. The free fraction in the circulation exerts the effects already described. Of note, dexamethasone does not bind to cortisol-binding globulin; therefore, almost 100% of it remains in the free, active form.

Cortisol and cortisone are used only for replacement therapy for adrenal insufficiency; they have a greater mineralocorticoid than glucocorticoid activity. Prednisone, prednisolone, and methylprednisolone are the glucocorticoids exogenously administered for their anti-inflammatory properties. They are used to treat allergic reactions (as elaborated in Chapter 3), inflammatory process (such as chronic sinusitis), autoimmune processes (e.g., collagen vascular diseases and acute sensorineural hearing loss), and vasculitic syndromes (e.g., Wegener's granulomatosis and giant cell arteritis). Dexamethasone and betamethasone have maximal anti-inflammatory properties and are used for acute inflammatory crises, but their side effects preclude long-term usage.

The side effects of glucocorticoids are as diverse as their range of activity. As already described, their ability to suppress the immune system promotes infection. They also promote osteoporosis by disrupting the regulation of calcium metabolism. Long-term usage is associated with the development of Cushing's syndrome and diabetes mellitus as well as the development of peptic ulcers. They can produce hypogonadism in men and anovulation and oligomenorrhea in women. Acutely, they

Table 8–1

Substance	Relative Anti-inflammatory Potency	Equivalent Dosage (mg)	Na Retention	Plasma Half-Life (Hours)
Hydrocortisone	1	20	++	1.5
Prednisone	4	5	+	2
Methylprednisone	4	4	−	2–4
Triamcinolone	5	4	−	3–5
Dexamethasone	25	0.75	− −	4–6

Na, sodium.

can produce central nervous system (CNS) effects such as arousal and euphoria.

Exogenous glucocorticoid administration produces a host of potential side effects; the withdrawal of this exogenous administration also produces a similar range of side effects because the HPA axis has been down-regulated and the body cannot produce enough endogenous glucocorticoid. To prevent the potentially serious morbidity and mortality (tiredness, nausea, vomiting, and hypotension) from too rapid a withdrawal from exogenous glucocorticoid administration, several precautions are taken. Some centers have taken to administering exogenous glucocorticoid on an alternating day regimen in the hopes of maintaining the anti-inflammatory responses while preventing the down-regulation of the HPA axis. Dosage patterns vary greatly; a common anti-inflammatory dosage for collagen vascular disorders or allergic reactions is 1 mg/kg/day of prednisone or the equivalent dose of another substance. The common practice is to taper the dosage of glucocorticoids if the equivalent of 5 mg of prednisone has been given for 2 weeks or more (**Table 8–1**). The short adrenocorticotropic hormone (ACTH) test is used to assess the recovery of the HPA axis. In this test, ACTH is given intravenously, and plasma cortisol levels are measured at 30- and 60-minute intervals after this administration. A normal HPA axis will allow for a greater than 20 μg/dL increase in the serum cortisol level.

NONSTEROIDAL ANTI-INFLAMMATORY DRUGS

Nonsteroidal anti-inflammatory drugs (NSAIDs) are weak organic acids that exert their effects by inhibiting the enzyme cyclooxygenase, thereby inhibiting the pathway from arachidonic acid (found in cell membranes) to various prostaglandins. They also exert their effects by interacting with the various other products of arachidonic acid metabolism; namely, leukotrienes. NSAIDs inhibit the many processes regulated by prostaglandins; they reduce fever by inhibiting prostaglandin-E_2 synthesis in the hypothalamic region and by blocking the hypothalamic response to interleukin-1 (IL-1), affect pain control by affecting prostaglandin synthesis at local sites, and also affect vascular permeability, the degree of leukocyte infiltration, and therefore the inflammatory response.

Of the NSAIDs, aspirin (acetylsalicylate) is perhaps the most commonly used. It is absorbed and hydrolyzed to acetic acid and salicylate in the tissue and serum and then is bound to albumin. It is excreted in the urine; this process of excretion is augmented by alkalinizing the urine, which is done in cases of overdosage. It acts by permanently acetylating an active site on the enzyme cyclooxygenase. It is effective as an anti-inflammatory, antipyretic, or analgesic agent. It is also an effective inhibitor of platelet aggregation. Its side effects include allergic reactions, gastrointestinal bleeding, coagulation disorders, renal abnormalities, and "salicylism"—tinnitus, hearing loss, and vertigo. Allergic reactions, especially in the face of nasal polyposis and asthma, can be severe and include bronchoconstriction and shock. The phenomenon of tinnitus derives from effects on the stria vascularis and is reversible upon discontinuation of the agent. The coagulation disorders engendered by aspirin require the regeneration of platelet cyclooxygenase for reversal; this process has a half-life of 7 to 10 days. Aspirin has also been associated with Reye's syndrome.

Because it has few side effects, acetaminophen is widely used as an NSAID. It is effective for local pain control and as an antipyretic; it has relatively little anti-inflammatory action. Ibuprofen (the active ingredient in Advil and Nuprin) is also superior in attaining pain control but is not an active anti-inflammatory agent. Other NSAIDs such as indomethacin have greater anti-inflammatory properties as compared with their ability to control pain or fever. Indomethacin, however, has a greater number of side effects, including the possibility of developing gastrointestinal hemorrhage and pancreatitis.

THE PHARMACOLOGY OF ANALGESICS

> I esteem it the office of the physician not only to restore health, but to mitigate pain and dolors.
>
> Sir Francis Bacon

Early humans attempted to relieve pain by entreaties to nature, sun or moon, and a variety of deities. Unfortunately, death was often the only effective surcease of pain. Homer, in the *Odyssey*, notes that Helen, the daughter of Zeus, prepared a drug (possibly opium), dissolved in wine, to ward off pain. In the 12th century Nicolas of Salerno suggested the use of the soporific sponge infused with a melange of possibly active and probably toxic substances whose inhalation not only could relieve pain but also could induce anesthesia. Also, it was reversible, with the judicious application of juice from the root of fennel. As unlikely as this magic sponge was in relieving pain, it does indicate some knowledge of the magnificent pain killer opium. By the 13th century opium, with a few fellow travelers, was used (undoubtedly with some success) to abate pain. Shakespeare, in *Romeo and Juliet, Othello,* and *Antony and Cleopatra,* mentions in each play a potent potion able to induce sleep and relieve pain. Knowledge of the beneficial effects of opium thus existed in crude form for many years, but it was not until 1806 that Friedrich Serturner isolated the active component of opium, which he named morphium (later morphine) after the Greek god of dreams, Morpheus. Although morphine has been followed by several effective pain relievers, none has yet to prove of greater value and reliability than the discovery of an early 19th-century Westphalian chemist.

Morphine

Opium is obtained from the seedpod of the poppy (*Papaver somniferum*). The dried juice can be broken down into its important component alkaloids, morphine, codeine, and papaverine. Chemically, morphine and codeine are phenanthrenes, and papaverine is a benzylisoquinoline. The chemical configuration of morphine can be manipulated to produce active compounds such as heroin, hydromorphone, oxymorphone, and codeine. None of these substances has demonstrated clinical superiority to morphine.

Morphine's pharmacological activity rests in its action as an agonist to receptors in the spinal and supraspinal CNS and bowel. In patients with pain, it produces both profound analgesia and drowsiness. In patients without pain, nausea and vomiting are common, and drowsiness may be exaggerated. In toxic doses morphine induces severe respiratory depression, which may be fatal. This depressant effect is due to morphine's ability to diminish markedly the respiratory stimulating action of carbon dioxide (CO_2). The obtunded patient will breathe on command but, without command, will simply stop breathing and asphyxiate.

Nausea and vomiting are more common in ambulatory patients than those who are supine. There is direct stimulation of the chemoreceptor trigger zone within the medulla of the brain and an increase in vestibular sensitivity. Although this unpleasant aspect of morphine has received significant attention, it should be understood that all agonists share the unpleasantry and that in equianalgesic doses, morphine is no greater a culprit than other commonly used analgesics.

In the supine patient, morphine has little effect on blood pressure and pulse, and cardiac inotropy is well maintained. However, morphine is a vasodilator, and, on sitting up, the subject may experience marked hypotension. Morphine causes the release of histamine, a potent vasodilator. The hypotensive effect is not due solely to histamine, but probably is an antiadrenergic effect, both centrally and peripherally. Because of its vasodilator effect and hypotensive consequences, morphine should be used with great care (if at all) in a hypovolemic patient.

Morphine also causes an increase in biliary pressure and may precipitate biliary colic. The drug inhibits the urinary voiding reflex and increases sphincter tone. Occasionally, urinary tract catheterization is necessary after the use of morphine.

Because the opioid analgesics are metabolized in the liver, liver disease with consequent decrease in metabolism will increase circulating levels of morphine. Renal disease may increase levels of the active metabolite morphine-6-glucuronide. Hepatic and renal disease thus demands caution on the part of the clinician when using opioids.

Morphine when injected intramuscularly or subcutaneously in a dose of 10 mg to an average adult is active for 4 to 5 hours and has a plasma half-life of 2 hours. Bioavailability after oral ingestion is very limited.

Codeine

The major advantage of codeine is its ready bioavailability when taken by mouth due to minimal first-pass metabolism in the liver. The drug has little analgesic ability per se, but is metabolized to morphine, which then exerts its effect. Codeine by mouth is 60% as effective as parenteral codeine. Because codeine is approximately one tenth as potent as morphine, 60 mg of oral codeine would be equal in analgesic potency to 3.5 mg of parenteral morphine ($60 \times 0.6/10$). The analgesic effect of the combination of codeine and aspirin exceeds the analgesic effect of either drug alone.

Meperidine

Meperidine is a phenylpiperidine whose chemical structure bears little obvious resemblance to that of morphine. However, its pharmacological properties are similar to those of morphine. Analgesia occurs 10 minutes after parenteral administration and peaks at approximately 1 hour. Meperidine is approximately one tenth as potent as morphine, with 100 mg of meperidine equivalent to 10 mg of morphine. Oral meperidine is approximately 50% as active as parenteral meperidine. The drug is metabolized in the liver to meperidinic acid and normeperidine. Normeperidine increases CNS excitation and can cause convulsions in patients with renal or hepatic insufficiency.

Agonist-Antagonist Analgesics

This group of drugs binds to the receptor and thus competes with other active drugs for the receptor site. Because they have little activity, they are either partial agonists or competitive antagonists. The group includes pentazocine, nalbuphine, butorphanol, and buprenorphine. Advocates of these drugs point to the "ceiling effect" on respiratory depression; that is, these drugs cause respiratory depression, but there is a peak beyond which further drug administration causes no increase in depression. However, respiratory depression is significant and can be hazardous. These drugs also have a capacity to counter the effects of the opioids and induce withdrawal symptoms in addicts or in those whose pain has been relieved by morphine or meperidine. They also have the ability to produce hallucinogenic effects with high doses.

Opioid Antagonists

Naloxone is a very effective opioid antagonist and can quickly counter the respiratory depressant effect of these drugs. Small doses (0.2–0.4 mg intravenously) exert an almost immediate action with little undesirable side effects. The sudden antagonism of the analgesic effect can cause marked increases in catecholamine levels, with significant increases in pulse and blood pressure.

Anesthetic Analgesics

Although morphine and meperidine have long been important components of the anesthesiologist's intraoperative armamentarium, these drugs have largely been supplanted in the operating room by a family of synthetic analgesics, including fentanyl, alfentanil, and sufentanil. Major advantages in this class of drugs include little or no histamine release and brief duration of action.

Use of Analgesics

Safe, effective pain relief is available for every patient and must be the clearly defined goal of the surgical and nursing staffs. Too often patients have suffered needlessly because of groundless fears of inducing addiction. Respiration must be monitored, and naloxone must be readily available. Recent advances in patient-controlled analgesia have placed patients in a decision-making role, thus providing them with the comfort of control.

SUMMARY OF ANALGESICS

Potent analgesic drugs are readily available to provide pain relief to surgical patients. Respiratory depression is an urgent consideration with any of these drugs but can be overcome with the administration of safe and effective antagonists. Patients need not, and should not, suffer.

THE PHARMACOLOGY OF LOCAL ANESTHETICS

Local anesthetics are safe, effective drugs, widely used when general anesthesia is inadvisable or not required by the nature of the operative procedure. The clinician must choose from a variety of available drugs. To make this choice, it is necessary to have knowledge of

- Chemistry of the local anesthetics
- Physiology of nerve transmission
- Pharmacology of local anesthetics
- Individual anesthetic drugs, their advantages and disadvantages
- Toxicity, prevention, recognition, and treatment

Chemistry

The clinically used local anesthetics have three chemical components:

- Lipophilic aromatic residue
- Hydrophilic amino group
- Linkage group, either ester or amide

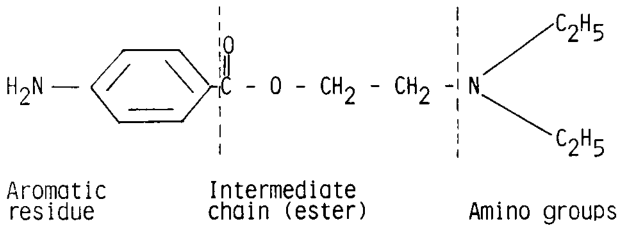


Figure 8–1 *Structural formula of procaine. (From Gotta AW, Donovan R, Sullivan CA. The pharmacology of local anesthetics. Ophthalmol Clin NA 1998;11(1). Reprinted with permission)*

An ester is a salt of an organic acid. An amine is a substitution product of ammonia (NH_3,), and an amide is a substitution product of NH_3, with substitution by an acid or oxidized radical.

Procaine is a typical ester local anesthetic (**Fig. 8–1**), with its aromatic residue linked by the ester intermediate chain to the terminal diethyl amino group. Tetracaine also has an ester link between aromatic and amino groups.

Lidocaine and bupivacaine differ from procaine and tetracaine in that their linking intermediate chain has an amide configuration (**Fig. 8–2**).

The local anesthetics may thus be divided into amide or ester anesthetics (**Tables 8–2, 8–3**) an important distinction because of the different metabolism of the two groups. The esters are metabolized by plasma (pseudo-) cholinesterase that is abundant in normal human blood. Rapid hydrolysis by pseudocholinesterase decreases the potential toxicity of ester-type local anesthetics, giving this class of drugs an inherent safety from the risks of excessive accumulation and toxicity. The amides are metabolized in the liver, a more prolonged process than rapid hydrolysis of the esters, although in blood-borne transit to the liver some of the anesthetic will find its way to the target organs of toxicity, the brain and heart.

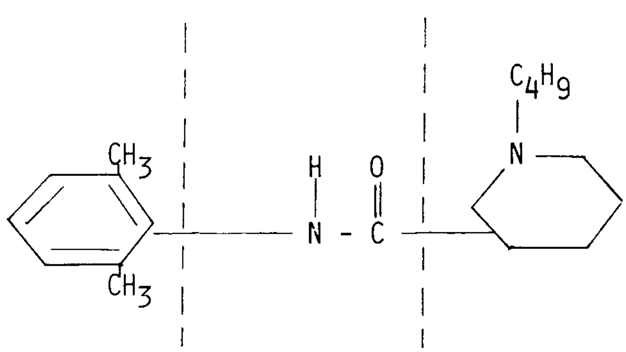


Figure 8–2 *Structural formula of bupivacaine. (From Gotta AW, Donovan R, Sullivan CA. The pharmacology of local anesthetics. Ophthalmol Clin NA 1998;11(1) Reprinted with permission.)*

TABLE 8–2 COMMONLY USED AMIDE LOCAL ANESTHETICS

Lidocaine (Xylocaine, lignocaine)
Bupivacaine (Marcaine, Sensorcaine)
Mepivacaine (Carbocaine)
Prilocaine (Citanest)
Ropivacaine (Naropin)

Liver disease may limit metabolism of either class of local anesthetic, but it is more likely to be clinically significant for the amides. The presence of an atypical pseudocholinesterase, or the presence of the enzyme in markedly diminished concentration, could cause prolonged metabolism, anesthetic effect, and increased risk of toxicity. Decreased levels of pseudocholinesterase may account for slower metabolism of procaine in newborn infants and in patients with liver disease. The commonly used muscle relaxant succinylcholine is also metabolized by pseudocholinesterase. A prolonged reaction to either local anesthetic or muscle relaxant should alert the clinician to the increased possibility of toxicity in the other class of drug.

The local anesthetics are tertiary amines and thus exist in the uncharged (cationic) or positively charged (protonated) form. The distribution relationship between the two forms is a function of the pH of the solution and the pKa of the anesthetic. The pKa is the pH at which a substance is 50% charged and 50% uncharged. This distribution is important because it is the uncharged form that crosses cell membranes (including neural), but it is the charged form that exerts the anesthetic effect. Nonionized local anesthetics are insoluble in water. They are soluble, however, when acidified (and thus protonated). When the aqueous solution of the protonated anesthetic is injected, it is buffered by tissue fluids, producing an uncharged form that crosses nerve membranes, where the drug molecules are again protonated by the lower pH within the axon. This protonated form is believed to produce neural blockade.

When a local anesthetic is injected into an area of decreased pH, the shift to the nonionized form will be

TABLE 8–3 COMMONLY USED ESTER LOCAL ANESTHETICS

Procaine
Chloroprocaine (Nesacaine)
Tetracaine (Pontocaine)
Cocaine
Hexylcaine (Cyclaine)

limited, resulting in a relative increase of the acidic form and thus decreasing the amount of mobile, nonionized form. Clinically, this results in lack of anesthesia. Cellulitis, with increased metabolism and cell breakdown, will produce an acidic environment, decreasing the effectiveness of local anesthetics.

Physiology of Nerve Transmission

Potassium is present in high concentrations in the axonal fluid and in low concentrations outside the nerve membrane. The nerve membrane is freely permeable to potassium. The energy expended to maintain the potassium gradient represents the resting membrane potential and is defined mathematically as membrane potential (millivolts) $= -61$ log intracellular K^+ concentration extracellular K^+ concentration. In contrast to potassium, sodium is present in high concentrations outside the resting nerve membrane but in low concentrations within the axon. The nerve membrane is not permeable to sodium, which can traverse the membrane only through specific lipoprotein pores.

Nerve transmission depends on opening the lipoprotein pores in the nerve membrane and the influx of sodium. To maintain ionic balance, potassium crosses the nerve membrane and becomes extraneural. Sodium is then actively pumped out of the cell, and potassium reenters and the resting state is then reestablished.

There are thus four possible mechanisms that might affect a pharmacological blockade of nerve transmission:

- Blocking the sodium channel at its external opening
- Blocking the sodium channel at its internal opening
- Incorporation of the local anesthetic into the nerve membrane, with consequent expansion of the volume of the membrane and physical compression of the sodium channel
- Blocking the regulator of the "gate" mechanism

The external opening of the sodium pore is blocked by the toxins saxitoxin and tetrodotoxin but is not the mechanism for the local anesthetics. There is no conclusive evidence that alteration in the gate mechanism is important.

Current opinion holds that the second and third mechanisms account for local anesthetic activity. According to this theory, a small part of the anesthetic action is due to incorporation of the uncharged local anesthetic molecule into the nerve with consequent membrane expansion. The majority of local anesthetic molecules pass through the sodium pores in their nonionized form. Within the axon they become protonated and then attach themselves to the axoplasmic opening of the pore, closing it to sodium flux and thus preventing nerve transmission.

Commonly Used Local Anesthetics

Cocaine

Cocaine is an ester-type drug, chemically similar to atropine and scopolamine. It is notable for its remarkable CNS effects and the marked euphoria it may produce.

Cocaine is a useful topical anesthetic. It is also a potent vasoconstrictor and helps to create a dry, bloodless field, with little edema. Vasoconstriction results from cocaine's ability to inhibit norepinephrine reuptake at the sympathetic nerve terminus. Although localized increases of norepinephrine occur, systemic levels will also increase and lead to tachycardia and hypertension, and can cause serious, even fatal, cardiac arrhythmias. Arrhythmias may occur at any time but are especially prone to occur with the concurrent use of inhaled anesthetics that sensitize the myocardium to catecholamines (e.g., halothane). Although the toxic concentration of cocaine is generally put at 200 mg in an adult (3 mg/kg), serious arrhythmias have been reported with concentrations as low as 50 mg.

Lidocaine

Lidocaine is an amide local anesthetic, suitable for a variety of clinical applications. It is an effective topical anesthetic at concentrations of 2 to 4%, and may be used for local infiltration at 0.5 to 1.0% and for nerve blocks in 1.0 to 2.0%. It has a rapid onset of action (usually 15 minutes or less) and is effective for 1½ to 2 hours without epinephrine and 2 to 4 hours when epinephrine (1:200,000) is added to the anesthetic.

Bupivacaine

Bupivacaine is an amide local anesthetic capable of providing prolonged block (up to 24 hours). Onset of block is slow, and it is not unusual to mix a rapid-acting local anesthetic (e.g., lidocaine) with bupivacaine to produce rapid onset and prolonged duration. Bupivacaine possesses a unique toxicity in that the premonitory CNS signs are often absent, and the first indication of toxicity is cardiac arrest. Bupivacaine-induced cardiac arrest may be very difficult to reverse, perhaps because the drug binds tightly to myocardial proteins.

Chloroprocaine

Chloroprocaine is an ester drug with rapid onset of action (less than 15 minutes), brief duration of action (30–45 minutes), and limited toxicity owing to rapid hydrolysis.

Ropivacaine

Ropivacaine is a recently introduced amide local anesthetic, a congener of mepivacaine and bupivacaine.

Ropivacaine and bupivacaine are equally potent, but ropivacaine (0.25–1.0%) appears to have a longer duration of action than bupivacaine and may be a vasoconstrictor over a wide range of concentrations. Studies suggest that ropivacaine has less cardiac toxicity than bupivacaine at equipotent concentrations but is more cardiotoxic than lidocaine.

Clinical Considerations

Physicians who use local anesthetics should familiarize themselves with at least one representative of each of three groups. Short-acting local anesthetics (20–45 minutes) include procaine and chloroprocaine. Intermediate-acting local anesthetics (60–120 minutes, longer with epinephrine) include lidocaine and mepivacaine. Long-acting anesthetics (400–450 minutes) include bupivacaine, etidocaine, and ropivacaine. The clinician can thus choose an anesthetic with a duration of action most closely approximating the duration of surgery. The longer acting local anesthetics may even be used for relatively brief surgical procedures to provide analgesia lasting well into the postoperative period.

The duration of block can be prolonged by the incorporation of a vasoconstrictor into the anesthetic mix. Epinephrine is most useful in a concentration of 1:200,000; stronger concentrations are not indicated. Epinephrine should be used with caution (if at all) in any patient where increased sympathetic activity might be deleterious (e.g., with heart disease, hypertension, or hyperthyroid). The drug should also be avoided in areas of end arterioles (e.g., fingers, nose, subcutaneously).

Toxicity

Misadventures with local anesthetics occur when the drugs are not effectively localized to the desired site of action but reach the target organs of toxicity, the brain and heart, in toxic concentrations. It is conventional to recommend a maximum dose of a particular local anesthetic, but the practice is arbitrary, unscientific, and misleading. Small doses of local anesthetic can cause a toxic reaction if they are injected directly into a blood vessel. Although the total amount of drug used plays a role in the development of toxicity, other factors include the location of injection and the incorporation of epinepherine. The vasoconstrictor thus serves to prolong the duration of block and minimize the risk of toxicity. Blood levels of local anesthetics are also a function of plasma protein binding of the drug, binding capacity that is variable and dependent on the patient's state of health. Alpha 1-acid glycoprotein (AAG) is the most important binding protein, and albumin is second in importance. AAG levels increase in chronic inflammatory illnesses, such as Crohn's disease, inflammatory arthritis, and chronic renal diseases with superimposed inflammatory disease. Theoretically, these disease processes would decrease the risk of local anesthetic toxicity by increasing AAG levels and decreasing the amount of unbound drug in the blood. **Table 8–4** gives the maximum doses but should serve only as a rough guide to the clinician.

Prevention of Toxicity

Appropriate premedication may serve to prevent or modify local anesthetic toxicity. The barbiturates are effective only at unacceptably large (and toxic) doses. However, the benzodiazepines are effective, probably because of the high density of benzodiazepine receptors in the limbic system of the brain, the primary site of toxicity.

Diazepam is water insoluble and will not produce an effective blood level when injected intramuscularly. It is soluble in propylene glycol, a venous irritant, causing pain on intravenous injection. Lorazepam is water miscible and can produce effective blood levels with intramuscular injection. Midazolam is water soluble and produces desired blood levels after either intramuscular or intravenous injection. The benzodiazepines are potent respiratory depressants and, in combination with narcotics, can produce significant hypotension. When benzodiazepines are used either as premedicant or for intraoperative sedation, the antagonist

Table 8–4 Suggested maximum doses of local anesthetics (mg/kg)

	Without Epinephrine	With Epinephrine	24-Hour Total
Lidocaine	300 (4.5)	500 (7)	
Bupivacaine	175 (2.5)	225 (3.2)	400
Chloroprocaine	800 (11)	1000 (14)	
Mepivacaine	400 (5.7)		1000
Ropivacaine	200 (2.8)	NA	750

NA, not applicable

Reproduced with permission from Gotta AW, Donovan R, Sullivan CA. The pharmacology of local anesthetics. Ophthalmol Clin NA 1998;11(1).

flumazenil must be readily available. Minimal effective concentrations of the local anesthetics should be used to decrease the total amount of anesthetic. For example, 1% lidocaine may be used for skin infiltration; 0.5% lidocaine is effective in subcutaneous areas.

Recognition of Toxicity

Local anesthetic toxicity is biphasic, beginning with alterations in the CNS and progressing to cardiac arrhythmias and cardiac arrest. Signs and symptoms of CNS toxicity include drowsiness (especially with lidocaine), personality changes, headache, tinnitus, tingling in lips or tongue, muscle tremors, and convulsions. Drowsiness is so common when lidocaine is used that it may be overlooked as a harbinger of worse things to come. The earliest indication of incipient toxicity is a change in personality. It is imperative that the anesthesiologist (or surgeon if no anesthesiologist is in attendance) maintain voice contact with the patient. Even if the patient has been premedicated with a benzodiazepine, a baseline can be established and variance noted. Deep sedation (as with, e.g., propofol) makes effective contact impossible. Tingling in lips or tongue is probably not a true sign of toxicity, but only a reflection of high concentrations of local anesthetics in richly perfused areas. These signs do indicate high blood levels of local anesthetic and serve to warn the clinician of troubles to come. Tremors begin in the small muscles of the hand or face, progressing to large muscle groups and frank convulsions. Recognition of the early signs of toxicity should prevent this occurrence.

Among the most commonly used local anesthetics bupivacaine causes the greatest depression of cardiac excitability and conduction, lidocaine the least and ropivacaine intermediate. Lidocaine overdose causes a marked reduction in myocardial contractility, bradycardia, hypotension, and respiratory depression. The electrocardiogram may demonstrate first-degree atrioventricular block and intraventricular conduction block. Bupivacaine overdose will cause ventricular tachycardia and fibrillation. The premonitory CNS signs may be lacking, and the first evidence of toxicity may be cardiac arrest. The toxicity of ropivacaine resembles that of lidocaine, but with periods of ventricular arrhythmia. Bupivacaine is approximately four times more potent than lidocaine as local anesthetic but is approximately nine times more toxic.

An interesting concept suggests that the lethality of bupivacaine is not caused by any direct toxic effect on the heart, but is actually a CNS toxicity caused by an imbalance between sympathetic and parasympathetic activity, causing hypertension and cardiac dysrhythmias. This suggests that appropriate therapy of bupivacaine toxicity might best be directed primarily at the brain, not the heart, and may explain the efficacy of benzodiazepines in treating bupivacaine overdose.

Management of Toxicity

Toxicity is best managed by avoidance, the use of suitable benzodiazepine premedication, administration of minimal effective concentrations of the drugs to avoid overdose, and the incorporation of epinephrine 1:200,000 when feasible into the anesthetic mixture. When large doses of local anesthetics are used, it is advisable to monitor the patient with electrocardiography to detect cardiac arrhythmias. A pulse oximeter is useful to detect hypoxia, and a capnograph may serve to detect hypoventilation, hypercarbia, and respiratory acidosis. Severe hypoxia increases the cardiotoxicity of bupivacaine.

Treatment of Toxicity

CNS hyperactivity may be treated with the intravenous administration of a benzodiazepine such as diazepam 0.1 mg/kg or preferably midazolam in 0.5 mg increments to a maximum of 5 mg. The anesthesiologist is more likely to use thiopental, a thiobarbiturate, in a dose of 1 to 2 mg/kg up to 4 to 5 mg/kg if necessary. These doses may be repeated if necessary if blood pressure is maintained and ventilation supported. Hypoxia and acidosis must be corrected.

If convulsions occur and persist despite appropriate therapy, muscle relaxants must be used to induce paralysis, and the patient then intubated and mechanical ventilation instituted. Hyperventilation will reduce $PaCO_2$ (partial pressure exerted by carbon dioxide dissolved in arterial plasma and red blood cell water) and increase the local anesthetic toxic threshold.

Hypotension is treated with intravenous fluids and positive inotropes (e.g., epinephrine, dopamine, or dobutamine). Bretylium (5 mg/kg) may be useful in refractory cardiac arrhythmias.

SUMMARY OF LOCAL ANESTHETICS

Local anesthetics are useful, safe drugs, applicable to a wide variety of surgical procedures. Adverse effects are rare and may be treated effectively if recognized early.

ANTIBIOTICS

As pharmaceutical companies keep pace with the development of bacteria resistant to antibiotics, the number of antibacterial agents continues to increase. Entirely new classes of antibiotics are also under development and may make their way onto the marketplace in the next few years. Bacteria develop resistance to antibiotics because of one simple reason: the use of antibiotics. Bacterial resistance is merely a survival technique and the result of evolutionary pressure on bacteria. Although much antibiotic use is justifiable and necessary to treat infections, excessive and unnecessary prescribing of antibiotics by physicians in many subspecialties, both for inpatients and outpatients, as well as the use of antibiotics in animal feed, contributes to needless bacterial resistance to antibiotics and the consequent occasional inability of physicians to treat bacterial infections successfully. The need for new antibiotics will continue for the foreseeable future.

All physicians should realize that it is impossible to eliminate bacteria from the world (and not even desirable to do so), and that use of antibiotics eventually always leads to bacterial resistance. A small fraction of the bacteria living in the body of a patient will become resistant to whatever antibiotic is given to that patient. If the number of bacteria in an infected site is relatively small, the infection will be cured. However, if the bacterial load is large, a superinfection by resistant bacteria may result. In either case, or if there is no bacterial infection at all, bacteria in the gastrointestinal tract or on the skin may acquire resistance. This is true even if the rate of acquiring resistance to a particular antibiotic is only one in a trillion (10^{-12}), because there are an estimated 10^{14} bacteria in the human body. If resistant bacteria multiply and cause infection, they may be difficult or even impossible to treat, and even if they do not cause infection, they may be spread to others. Therefore, physicians must use good medical judgment in deciding whether or not to treat a patient with antibiotics.

It is also important to note that cultures should be taken in the correct manner. Unless a particular species of bacteria is being sought, such as *Streptococcus pyogenes* from a throat culture, swab cultures should never be done. Even in an obviously infected area, swab cultures cannot distinguish the bacteria that are causing the infection from those that are simply on the surface. Instead, the infected area should be debrided or the infected body cavity should be entered in a sterile manner for cultures to be properly done.

In the subspecialty of otorhinolaryngology, these principles are most easily demonstrated with sinusitis and otitis, many cases of which are viral in origin. Indiscriminate use of antibiotics for these conditions, both in adults and children, undoubtedly contributes to resistance by many common bacteria. However, viral and bacterial forms of these diseases are impossible to distinguish on clinical grounds, and swab cultures of ears and sinuses are useless. Moreover, invasive techniques such as tympanocentesis and sinus puncture to make etiologic diagnoses are not justified on a routine basis. Therefore, clinicians are often forced to treat these conditions with antibacterials without proof of a bacterial cause.

Finally, it is important to note that even if a package insert states that a particular antibiotic is indicated for a specific condition, the phrase "due to susceptible bacteria" is always implied if not explicitly stated. Because of increased bacterial resistance, some antibiotics are no longer as useful for certain types of infections as they were when they were first marketed, and they should no longer be used empirically to treat these conditions. Conversely, just because an antibiotic does not have a particular indication does not mean that it will not work, and there are situations in which it is reasonable to use an antibiotic in an "off-label" manner, especially when multiresistant nosocomial bacteria are suspected or proven. Physicians should know the susceptibility patterns of common bacteria in their own local hospitals so that they can make a reasonable choice of empirical antibiotics for their hospitalized patients. The microbiology laboratories of most hospitals are glad to supply this information to their physicians.

Groups of antibiotics and individual agents are reviewed on the following pages, with some focus on their use for infections of the ears, nose, and throat, as well other uses that otorhinolaryngologists are likely to encounter. When brand names are commonly used for a specific antibiotic, they are supplied in parentheses after the generic names. For more detailed reviews of antibiotics, and for thorough discussions of their uses for other types of infections, a variety of texts may be consulted (see Suggested Readings). A multitude of pocket-sized handbooks (e.g., *The Medical Letter Handbook of Antimicrobial Therapy, The Sanford Guide to Antimicrobial Therapy*) have also been published that summarize antibiotic dosing, empirical selection by site of infection, dosages in patients with renal failure and in pediatric patients, achievable serum concentrations, side effects, and so on. Even clinicians who are familiar with antibiotics will find one or more of these handbooks useful for prescribing antibiotics in certain situations.

Sulfonamides and Trimethoprim

Sulfonamides were the first type of modern antibiotic used. The first scientific report of antibacterial activity described the use of sulfachrysoidine in the treatment of streptococcal infections in mice. Sulfanilamide (Prontosil) became the first antibiotic to be widely used in humans after scientists discovered that it was the active metabolite of sulfachrysoidine. After scientists realized that chemical modifications of the sulfanilamide molecule greatly expanded its antibacterial activity and reduced its side effects, thousands of sulfonamides were studied, and dozens were used to treat patients. Only a small fraction of these are used nowadays, but they are still valuable antibiotics.

Currently, the sulfonamide most frequently used for systemic administration is sulfamethoxazole, which is almost always used in a fixed combination with trimethoprim. This preparation is marketed under several brand names, such as Bactrim, Cotrim, and Septra, and is also available generically. The different brands are equivalent. Single-strength (SS) pills consist of 80 mg of trimethoprim and 400 mg of sulfamethoxazole, and double-strength (DS) pills contain 160 and 800 mg, respectively. It is useful for the treatment of acute otitis media and sinusitis, as well as for urinary tract infections, exacerbations of chronic bronchitis, and other conditions.

Another sulfonamide that is still used by pediatricians in some areas is a suspension of sulfisoxazole in a fixed combination with erythromycin, known as Pediazole. Each teaspoon (5 mL) contains 600 mg of sulfisoxazole and 200 mg of erythromycin. It is indicated only for otitis media due to susceptible strains of *Haemophilus influenzae*, which is too narrow an indication for routine empirical use because other species of bacteria that cause otitis media, especially *Streptococcus pneumoniae*, are frequently resistant.

All currently used sulfonamides consist of a benzene ring with an aminosulfonyl group on C_1 and an amino group on C_4. Addition of a heterocyclic aromatic ring on N_1 of the aminosulfonyl group, as in sulfamethoxazole, greatly increases antibacterial activity compared with the original sulfanilamide (**Fig. 8–3**).

Chemically, sulfonamides resemble para-aminobenzoic acid (PABA), and their mechanism of action is to inhibit the enzyme dihydropteroic acid synthetase, which catalyzes the incorporation of PABA into dihydropteroic acid to form dihydrofolic acid. The latter compound is important in the transfer of single carbon fragments in numerous biochemical pathways. Humans cannot synthesize dihydrofolic acid and must ingest folic acid in their diet or in vitamin supplements. Therefore, sulfonamides do not adversely affect humans in the same way that they harm bacteria.

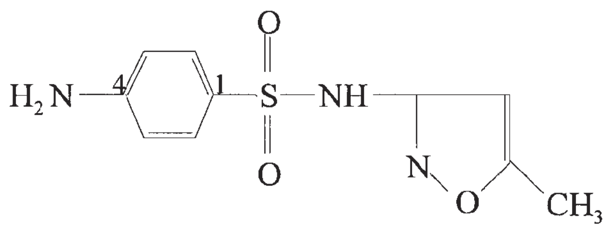


Figure 8–3 *Molecular structure of sulfamethoxazole.*

Bacteria become resistant to sulfonamides via overproduction of PABA to counteract the competitive inhibition by sulfonamides. This is accomplished by production of a mutant dihydropteroic acid synthetase that does not bind well to sulfonamides (alteration of the target site), or by decreased uptake of sulfonamides.

The spectrum of activity of sulfonamides includes streptococci (including some strains of *S. pneumoniae*), *H. influenzae*, and *Moraxella catarrhalis*. Strains of *S. pneumoniae* that are resistant to penicillin are also frequently resistant to other classes of antibiotics as well, including sulfonamides, and these strains account for 40% of isolates in some areas of the United States. Outpatient-type enteric Gram-negative bacilli, including *Escherichia coli*, *Klebsiella* spp., *Enterobacter* spp., *Morganella morganii,* and *Proteus* spp., are also usually susceptible to sulfonamides. Sulfonamides should not be given for the treatment of sore throats because they do not eradicate group A streptococci (*S. pyogenes*) from the tonsils as well as penicillin does.

Sulfonamides are fairly well absorbed when taken by mouth, and the trimethoprim-sulfamethoxazole combination needs to be dosed only twice a day for the treatment of most types of infections. Once absorbed into the bloodstream, they distribute fairly well to most body tissues, including the CNS. Metabolism is by acetylation and glucuronidation, which occur in hepatocytes. Glucuronidation increases solubility in water so that excretion, which is via glomerular filtration, can occur.

The most common side effects are rashes due to allergic hypersensitivity; most often, this is an erythematous rash that disappears after the sulfonamide is stopped. Occasionally, Stevens-Johnson syndrome (erythema multiforme major), a severe, blistering, and desquamating mucosal rash, may occur. This is sometimes preceded by other types of rashes, but not always. Sulfonamides should not be given late in pregnancy or to newborns because they compete with bilirubin for the negligible amounts of glucuronyl transferase in the bloodstream of newborns. If the bilirubin cannot be conjugated, kernicterus may result. Other rare but serious side effects include fulminant hepatic necrosis and agranulocytosis.

Trimethoprim was developed for use with sulfamethoxazole because it has a similar mechanism of action, spectrum of activity, and pharmacokinetics. It is a 2,4-diaminopyrimidine and is not chemically related to the sulfonamide group. The only other chemically related antibiotic is pyrimethamine, which is an antiparasitic agent. (Trimethoprim binds very well to bacterial dihydrofolate reductase, and pyrimethamine binds very well to parasitic dihydrofolate reductase; neither one binds very well to the human enzyme.) Dihydrofolate reductase reduces dihydrofolic acid to tetrahydrofolic acid. Because this biochemical reaction directly follows the one that is inhibited by sulfonamides, it was reasoned that combining trimethoprim with sulfamethoxazole would have a synergistic action, and this has been borne out by in vitro testing. (In microbiology, synergy means that bacterial killing actually obtained as a result of combining two antibiotics is more than the killing expected by simply adding their individual effects.)

A 1:5 fixed ratio of trimethoprim to sulfamethoxazole was selected because this combination yields a concentration of ~1 to 20 in serum. (Sulfamethoxazole is absorbed ~4 times as well as trimethoprim.) Because trimethoprim is ~20 times as active as sulfamethoxazole on a milligram per milligram basis, this yields equal activities of both.

Distribution of trimethoprim into various body tissues and fluids is also very good, with therapeutic levels found in many areas. About 60% of a dose is excreted unchanged by renal tubular secretion in urine within 24 hours, with the remainder excreted in bile or undergoing oxidation or hydroxylation in hepatocytes to inactive metabolites. Side effects are also similar in quality to those due to sulfonamides, although rashes occur much less frequently.

Penicillins

Although Alexander Fleming noted in 1928 that *Staphylococcus aureus* was unable to grow near colonies of *Penicillium notatum*, the full importance of this finding was not understood, until systematic studies of penicillins were undertaken several years later. It was not until 1941 that penicillin G was first used in humans, and its development would have been further delayed if it had not been needed to fight battlefield infections during World War II. The penicillins, cephalosporins, and carbapenems are the three classes of β-lactam antibiotics.

The earliest forms of penicillin were a mixture of different natural penicillins that were designated by letters. The most active of these compounds was penicillin G (benzylpenicillin), which is still used for parenteral administration. Penicillin G is available in two crystalline

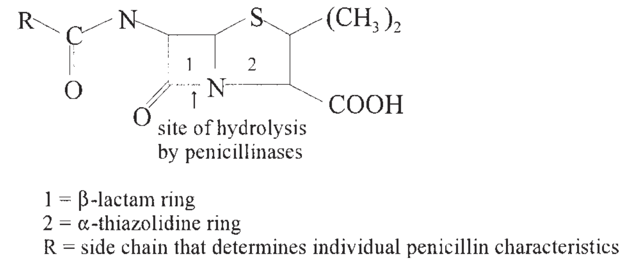


Figure 8–4 *Penicillin nucleus.*

forms (as sodium and potassium salts), as a less soluble procaine salt, and as an even less soluble benzathine (dibenzyl ethylene diamine) salt. Penicillin V (phenoxymethylpenicillin) is used for oral administration because it is more resistant to hydrolysis mediated by stomach acid.

Even though the initial antibacterial property of penicillin was observed on *S. aureus*, nearly all strains of these bacteria are now resistant to penicillin G and penicillin V due to the production of a penicillinase that hydrolyzes the penicillin molecule (**Fig. 8–4**). Resistance to penicillin G and penicillin V by *S. aureus* was first noted in the 1950s. Both these forms of penicillin have remained active against most streptococci, especially anaerobic streptococci found in the mouth and β-hemolytic streptococci. However, in the past 10 years, many strains of *S. pneumoniae* have become intermediately or highly resistant to penicillin, with rates of resistance approaching 40% in some areas of the United States and even higher rates elsewhere.

Because the individual penicillins in the initial natural mixtures differed only in their side chain, or "R group," scientists reasoned that by either modifying the side chain or putting a synthetic side chain onto the penicillin nucleus, they might be able to increase the antibacterial activity of the natural penicillins. There are several groups of these semisynthetic penicillins that have activity against staphylococci, enteric Gram-negative bacilli, and other bacteria. These are outlined as follows.

Penicillinase-resistant (antistaphylococcal) penicillins: methicillin, oxacillin, nafcillin, cloxacillin, and dicloxacillin. These all have very large R groups that stereochemically prevent penicillinases produced by staphylococci from getting close to the β-lactam ring and hydrolyzing it. Although methicillin is no longer used clinically, it is still employed by microbiology laboratories for susceptibility testing of staphylococci to all penicillins, cephalosporins, and carbapenems. Oxacillin and nafcillin are for intravenous administration; cloxacillin and dicloxacillin are for oral administration. Penicillins in this group are also active against streptococci, but to preserve their effectiveness, they should be used only for the treatment of infections due to staphylococci that are susceptible to methicillin. They are inactive against enterococci and Gram-negative bacilli.

Aminopenicillins: ampicillin and amoxicillin. The amino group on the carbon next to the benzyl side chain allows better entry than that of penicillin G and penicillin V into Gram-negative bacteria. These were originally called "broad-spectrum" penicillins because they had activity against enteric Gram-negative bacilli (especially *E. coli* and *Proteus mirabilis*) and *H. influenzae*, as well as streptococci and enterococci. Many strains of Gram-negative bacilli are now resistant to ampicillin and amoxicillin, making their empiric use unwarranted for serious infections. However, if culture results show susceptibility, these can still be used. Amoxicillin is used extremely often by pediatricians for the treatment of otitis media because the oral liquid formulation tastes good, making its administration easy in children. The good clinical results suggest that many cases of otitis media do not need treatment with antibiotics because a large proportion are caused either by bacteria that are amoxicillin resistant or by viruses. It is available only in oral formulations (pills and liquid). Ampicillin is available in both oral and intravenous formulations.

Carboxypenicillins: carbenicillin (Geocillin) and ticarcillin (Ticar). The carboxyl group on the carbon next to the benzyl side chain facilitates entry into *Pseudomonas aeruginosa*, making these antibiotics the first that were active against this species. They also have good activity against other Gram-negative bacilli, in addition to activity against streptococci that are susceptible to penicillin and staphylococci that are susceptible to methicillin. However, neither of these is used frequently nowadays. The absorption of oral carbenicillin is very limited, so it is useful only for the treatment of urinary tract infections, and its intravenous formulation is not as active as penicillins in the ureido group. Ticarcillin, which is available only intravenously, is now almost always used in the ticarcillin-clavulanic acid form (see below).

Ureidopenicillins: mezlocillin (Mezlin) and piperacillin (Pipracil). These have a ureido group, which facilitates their entry into *P. aeruginosa*. They also have good activity against penicillin-susceptible streptococci and enterococci, methicillin-susceptible staphylococci, and enteric Gram-negative bacilli. They are available only in intravenous formulations.

β-lactam–β-lactamase inhibitor combinations: ticarcillin-clavulanic acid (Timentin; intravenous), amoxicillin-clavulanic acid (Augmentin; oral), ampicillin-sulbactam (Unasyn; intravenous), and piperacillin-tazobactam (Zosyn; intravenous). The β-lactamase inhibitor component of these preparations chemically resembles the β-lactam ring of penicillin and binds to β-lactamases that are produced by bacteria. Because the β-lactamase inhibitors have very little antibiotic activity themselves, their function is to prevent the hydrolysis of the penicillin component via their binding to β-lactamases. Addition of the β-lactamase inhibitor component significantly enhances the antibiotic spectrum of the penicillin component against bacteria that are resistant to penicillins via production of β-lactamases. Generally, only β-lactamases that are encoded by plasmids bind to β-lactamase inhibitors. Therefore, some bacteria, most notably *P. aeruginosa*, which produce chromosomally encoded β-lactamases, are no more susceptible to piperacillin than piperacillin-tazobactam. Staphylococci, enteric Gram-negative bacilli such as *E. coli*, *Klebsiella pneumoniae*, and *Proteus* spp., and Gram-negative anaerobes such as *Bacteroides fragilis*, all of which produce β-lactams that are encoded by plasmids, are more likely to be susceptible to the combinations than to the corresponding penicillin without the β-lactamase inhibitor. However, strains of penicillin-resistant *S. pneumoniae* are no more likely to be susceptible to β-lactam–β-lactamase inhibitor combinations than they are to the corresponding β-lactam alone because their resistance to β-lactams is mediated by altered penicillin-binding proteins.

All penicillins have the same mechanism of action. Their chemical structure is similar to that of D-alanyl-D-alanine in the polypeptide portion of peptidoglycan in the bacterial cell wall, which is hydrolyzed by transpeptidases that act as penicillin-binding proteins. All bacteria have several different types of penicillin-binding proteins. After penicillins penetrate the bacterial envelope, they attach to penicillin-binding proteins, which starts a series of events involving accumulation of bacterial cell wall precursors that cannot be incorporated into the growing cell wall, activation of murein hydrolases that are normally involved in the regulation of growing cell walls, and finally bacterial death.

Resistance to penicillins by staphylococci, enteric Gram-negative bacilli, anaerobic Gram-negative bacilli, and some other bacteria (e.g., *Neisseria gonorrhoeae*) is generally due to production of β-lactamases. This is the most clinically important type of bacterial resistance because it is the most common type; in the case of β-lactamases encoded on plasmids, it can be passed from bacterium to bacterium (sometimes even across species), and no amount of penicillin can overcome this type of bacterial resistance. Hundreds of different types of β-lactamases have been discovered and characterized, but the individual types generally are not important to clinicians. A β-lactamase that is active only against penicillins is sometimes termed a penicillinase, and one that is active only against cephalosporins is sometimes called a cephalosporinase.

A second mechanism of resistance, which occurs only in Gram-negative bacilli, is a permeability barrier of the outer membrane. This is accomplished by the production of mutant porins, which are the channels in the outer membrane through which penicillins gain access to the cell membrane.

In the past decade, a third mechanism of resistance has become clinically important, that of the production of altered penicillin-binding proteins, which bind penicillin poorly. *S. pneumoniae*, which in the United States had been universally susceptible to penicillins prior to the 1980s, is now resistant at rates that make the empirical use of penicillins no longer justifiable when serious pneumococcal infections are suspected. The addition of a β-lactamase inhibitor does nothing to overcome this resistance. Third-generation cephalosporins are still useful when this type of resistance is present at an intermediate level because they still bind well enough to the penicillin-binding proteins to have antibacterial activity. However, when high-level penicillin resistance is present, even cephalosporins do not bind to the penicillin-binding proteins, and a macrolide, a quinolone with good streptococcal activity, or vancomycin should be used.

The availability of oral and intravenous formulations of penicillins has been described with each individual type of penicillin. Absorption of oral formulations depends on the particular penicillin; for example, only ~60% of an oral dose of penicillin V, 40% of a dose of ampicillin, 75% of a dose of amoxicillin, and 50% of a dose of dicloxacillin are absorbed from the stomach.

In patients with normal renal function, the half-life of most penicillins is relatively short, ranging from as little as 30 minutes for penicillin G and the penicillinase-resistant penicillins to as long as 1.3 hours for piperacillin.

After absorption or intravenous administration, penicillins generally achieve reasonably high tissue and fluid concentrations, with the notable exceptions of the prostate, joint cavities, and cerebrospinal fluid (CSF). However, in the presence of meningeal inflammation, CSF concentrations reach ~15% of simultaneous serum concentrations and are adequate to treat meningitis due to *Neisseria meningitides* and susceptible strains of *S. pneumoniae*.

Most penicillins are not metabolized to a significant degree. In the case of penicillin G, ~20% of a dose is metabolized by hepatic enzymes. Excretion of penicillins is nearly all renal, and active renal tubular secretion by an organic acid transport system accounts for their relatively short serum half-lives. Probenecid, which competes for penicillins in this transport system, can be given concurrently to prolong the half-life and raise serum levels.

The most important side effects of penicillins are hypersensitivity reactions. Immediate reactions occur within minutes of a dose and in their most severe form consist of hives, urticaria, flushing, wheezing, laryngeal edema, and shock. This is an immunoglobulin E (IgE)–mediated allergic reaction, similar to other forms of anaphylaxis, and occurs with a frequency of 0.004 to 0.4% of courses of treatment. Anaphylaxis to penicillin is more common in patients with a history of atopy, asthma, and other allergies.

Delayed or late allergic reactions, which are mediated by IgM or IgG antibodies, are more common hypersensitivity reactions, occurring in the range of 4 to 8%, and usually are manifested as morbilliform or macular rashes, interstitial nephritis, hemolytic anemia, or serum sickness. There is no way to predict whether a patient who is allergic to one penicillin will be allergic to another. Similarly, the chance that a person with a penicillin allergy is also allergic to cephalosporins or carbapenems is in the range of 1 to 5%, but there is no way to predict a cross-reaction in a particular patient. Because of this, and because other classes of antibiotics are available to treat infections, it is prudent to use a different type of antibiotic if a patient has an immediate hypersensitivity reaction. However, if the type of allergy is delayed hypersensitivity and a penicillin, cephalosporin, or carbapenem is needed, they may be used.

A less common but important side effect of penicillins is seizures if they are given at high doses, especially to patients with mass lesions in the brain, renal insufficiency, or hyponatremia.

Cephalosporins

After the penicillins, cephalosporins were the second class of β-lactam antibiotics developed. The first cephalosporin was isolated by Giuseppe Brotzu from a sewage outlet on the island of Sardinia after he noted that the seawater in the vicinity of the outlet was intermittently free of bacteria. The source of the cephalosporin was a fungus named *Cephalosporium acremonium*, which was in the water.

Because of their relatively broad antibacterial activity and lack of toxicity, more than 30 different cephalosporins have been marketed in the United States over the last 4 decades. However, many are very similar and therefore can be grouped together. As with other

TABLE 8–5 COMMONLY USED CEPHALOSPORINS

First Generation[a]	Second Generation	Third Generation	Fourth Generation
Oral cefadroxil (Duricef), cephalexin (Keflex), cephradine (Anspor, Velosef)	No activity versus *Bacteroides* spp. cefuroxime[b] (Zinacef, Kefurox, Ceftin) cefaclor (Ceclor)	Intravenous, no *Pseudomonas* coverage but good activity versus Gram-positive cocci cefotaxime (Claforan), ceftriaxone (Rocephin), ceftizoxime (Cefizox)	Good activity against both *Pseudomonas* spp. and Gram-positive cocci cefepime (Maxipime)
Intravenous cefazolin (Kefzol, Ancef)	With activity versus *Bacteroides* spp. cefotetan (Cefotan), cefoxitin (Mefoxin)	Intravenous, good *Pseudomonas* coverage but poor activity versus Gram-positive cocci ceftazidime (Fortaz, Tazicef, Tazidime), cefoperazone (Cefobid)	
		Oral cefixime (Suprax), cefprozil (Cefzil), cefpodoxime (Vantin)	

[a] Cephalothin, although no longer available in the United States for use in patients, is the representative first-generation cephalosporin used for in vitro susceptibility testing by many microbiology laboratories (the same situation as for methicillin).
[b] Cefuroxime is available both intravenously and orally; the other second-generation cephalosporins are available for intravenous administration only.

antibiotics, large marketing efforts by pharmaceutical companies frequently make distinctions without differences in their efforts to promote one cephalosporin over another. Distinguishing genuine variations between cephalosporins is even more difficult for physicians who are not experts in antibiotics, microbiology, or infectious diseases.

The most widely used classification system for cephalosporins categorizes them as first, second, third, and fourth generation. **Table 8–5** is meant to be used as a guide and does not list all available cephalosporins.

Clinicians should be aware that for nearly all purposes, cephalosporins that are grouped together in **Table 8–5** are the same. The price that a patient pays to a retail pharmacy, or the cost to the hospital pharmacy from the wholesaler, or formulary availability (either a hospital or insurance formulary) should determine the selection of a cephalosporin within each grouping in **Table 8–5**.

The antibacterial activity of cephalosporins requires the 7-aminocephalosporanic acid nucleus. The spectra of activity of individual cephalosporins are determined by the structures of their side chains. The chemical structure of the cephalosporin nucleus is shown in **Fig. 8–5**.

Cephalosporins have the same mechanism of action as penicillins. By binding to penicillin-binding proteins, which regulate bacterial cell wall synthesis, they induce several changes in bacterial metabolism, including the production of autolysins, ultimately resulting in cellular lysis. They bind to different penicillin-binding proteins in different proportions than do the penicillins.

The most common mechanism of bacterial resistance to cephalosporins is production of β-lactamases, which result in the hydrolysis of the β-lactam ring. Many different types of β-lactamases have been identified. Some are more active than others; some are constitutively produced, whereas others are inducible; and some are encoded by chromosomal genes, whereas others are produced by genes that are located on plasmids. Additionally, certain cephalosporins are more resistant to β-lactamases than others, whereas some are better inducers of β-lactamases than others. Therefore, one cephalosporin may be inactive against a certain strain

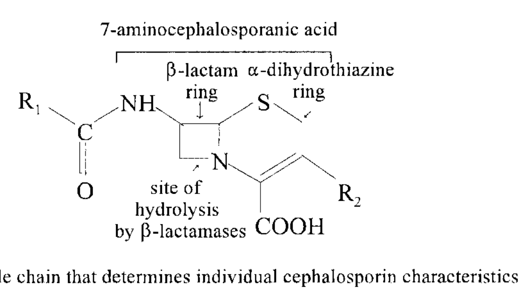


Figure 8–5 *Cephalosporin nucleus.*

of bacteria that produces a certain β-lactamase, whereas another may be active. Although closely related cephalosporins tend to have similar antibacterial activity, there are important differences that require testing of the in vitro activity of several cephalosporins (usually one for each generation) against bacteria from clinical samples by a microbiology laboratory.

Additional mechanisms of resistance include decreased bacterial permeability to cephalosporin molecules (by aerobic Gram-negative bacilli) and bacterial production of altered penicillin-binding proteins that bind poorly to cephalosporins (usually by Gram-positive cocci).

The side chains of individual cephalosporins determine the amount of their uptake into bacterial cells as well as their binding affinities to the various penicillin-binding proteins, resulting in different spectra of activity.

First-generation cephalosporins are active against all streptococci and methicillin-susceptible staphylococci, and against most community-acquired strains of *E. coli*, *Klebsiella* spp., and *Proteus* spp.

Second-generation cephalosporins have slightly less activity against streptococci and methicillin-susceptible staphylococci than first-generation cephalosporins (but still enough to be useful), but more activity against enteric Gram-negative bacilli. Cefuroxime (Zinacef, Kefurox, Ceftin) has excellent additional activity against *H. influenzae* (making it useful for the treatment of acute sinus, ear, and respiratory pathogens), and cefotetan (Cefotan) and cefoxitin (Mefoxin) have excellent additional activity against *B. fragilis* and other anaerobic bacteria.

The intravenous third-generation cephalosporins cefotaxime (Claforan), ceftriaxone (Rocephin), and ceftizoxime (Cefizox) have additional activity against many nosocomial Gram-negative bacilli, including *Enterobacter* spp., *Citrobacter* spp., and *Serratia marcescens*. Ceftazidime (Fortaz, Tazicef, Tazidime) also has activity versus *Pseudomonas* spp., but is significantly less active against Gram-positive cocci than most other cephalosporins. The oral third-generation cephalosporins have activity against penicillin-susceptible *S. pneumoniae*, *H. influenzae*, and *M. catarrhalis*, making them attractive to use for the treatment of acute respiratory, ear, and sinus infections in children.

Cefepime (Maxipeme), the first fourth-generation (or extended spectrum) cephalosporin, has Gram-negative activity similar to that of ceftazidime (i.e., enteric Gram-negative bacilli including nosocomial strains, and *Pseudomonas* spp.), and also has good activity against Gram-positive cocci.

All cephalosporins are inactive against enterococci, methicillin-resistant staphylococci, and *Listeria monocytogenes* because the penicillin-binding proteins produced by these bacteria do not bind cephalosporins.

Most cephalosporins can be administered either orally or intravenously, but not both. The only exception to this generalization is cefuroxime. The methods of administration of the individual cephalosporins are outlined in the classification system in **Table 8–5**.

Although most cephalosporins achieve concentrations in many body tissues and fluids that are adequate to treat infections at these sites, only ceftriaxone, cefotaxime, ceftizoxime, and ceftazidime achieve adequate concentrations in CSF and are considered reasonable treatment for bacterial meningitis. (These are all third-generation cephalosporins. Cefepime also enters the subarachnoid space, but there are not enough clinical data to warrant its use in treating meningitis.) Ceftriaxone also achieves relatively high biliary concentrations.

The cephalosporins are excreted unmetabolized, except for cefotaxime, which is deacetylated, and this metabolite also has antibacterial activity.

The usual mechanism of excretion is via glomerular filtration and tubular secretion, just as for penicillins. Because elimination is exclusively via the kidneys, dosages of cephalosporins must be reduced when they are given to patients with renal failure, or else serum concentrations become excessively high and may lead to seizures. Renal tubular reabsorption can be blocked with probenecid to prolong the serum half-life, as with the penicillins. The only important exception to exclusively renal elimination is ceftriaxone, which is eliminated via bile into feces as well.

As a group, the cephalosporins are relatively safe antibiotics. Side effects are very similar to those caused by the penicillins, in both type and frequency. The most common are allergic hypersensitivity reactions, which range from maculopapular rashes (mild and relatively common) to anaphylaxis (severe but relatively rare). About 1 to 5% of patients who are allergic to penicillin are also allergic to cephalosporins.

Cephalosporins that have a methylthiotetrazolium side chain have been associated with significant gastrointestinal bleeding due to the inhibition of post-translational carboxylation of glutamic acid residues in vitamin K–dependent clotting factors. These cephalosporins are cefamandole (Mandol), cefmetazole (Zefazone), cefotetan (Cefotan), cefoperazone (Cefobid), and moxolactam (Moxam). However, except for cefotetan, nowadays these are rarely used, no longer marketed, or no longer even made.

Other occasional side effects are thrombocytopenia, interstitial nephritis, fever, serum sickness, and *Clostridium difficile* colitis.

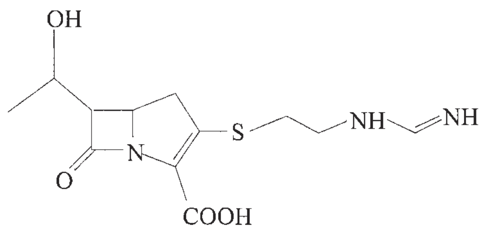


Figure 8–6 *Structure of imipenem.*

Carbapenems

These are also β-lactam antibiotics, but their side chains are in a *trans* configuration in relation to the β-lactam ring, which makes them much more resistant to hydrolysis by β-lactamases and therefore much more active against a wide variety of bacteria. Compared with the structure of penicillins and cephalosporins, there are additional differences to the α-thiazolidine ring and the side chain (**Fig. 8–6**). The three carbapenems currently available are imipenem (Primaxin), meropenem (Merrem), and ertapenem (Invanz).

The mechanisms of action of carbapenems are the same as that of penicillins and cephalosporins. The mechanisms of bacterial resistance are also the same, except that the *trans* ring configuration makes them poor substrates for nearly all β-lactamases, so both inherent and acquired resistance is much less common. When resistance occurs, it is often due to decreased permeability of the bacterial cell, altered penicillin-binding proteins, or production of a particular β-lactamase that has the ability to hydrolyze carbapenems.

The spectrum of activity of carbapenems includes streptococci, methicillin-susceptible staphylococci, ampicillin-susceptible enterococci, nearly all Gram-negative bacilli, and anaerobes. Notable exceptions are *Stenotrophomonas maltophilia*, *Burkholderia cepacia*, ampicillin-resistant enterococci, and methicillin-resistant staphylococci. Ertapenem is inactive against all *Enterococcus* spp. and *Pseudomonas* spp. Carbapenems should be used only for hospital-acquired, severe, or mixed infections, and not for routine empirical treatment of a patient requiring hospitalization of a mild or moderate infection. They cannot be given orally because they are hydrolyzed by stomach acid.

After administration, they enter all body tissues and fluids, except urine for imipenem. Upon reaching the kidneys, imipenem is hydrolyzed by a dehydropeptidase in the brush border of the proximal renal tubule cells. Its metabolites are also somewhat nephrotoxic. Therefore, commercially available imipenem is combined with a compound called cilastatin, which is a competitive inhibitor of the dehydropeptidase. This prevents the hydrolysis of imipenem, allowing it to reach therapeutic concentrations in urine, and also eliminates its nephrotoxicity. Meropenem and ertapenem are not metabolized and do not need to be given with cilastatin. Excretion is similar to that of most penicillins and cephalosporins and is via the kidneys.

Side effects are also similar to those of penicillins and cephalosporins (including ~1–5% hypersensitivity in patients who are allergic to penicillins and cephalosporins). Carbapenems can cause seizures when given at high doses, especially to patients with renal insufficiency or underlying abnormalities of the CNS. Unless there are no alternatives, one carbapenem should not be substituted for another if this happens.

Monobactams

Aztreonam (Azactam) is the only member of this class of antibiotics that is currently available. It is a monocyclic lactam ring to which a side chain identical to that of ceftazidime is attached (**Fig. 8–7**). Because it has only one ring, it is not a β-lactam antibiotic.

Its mechanism of action is similar to that of penicillins, cephalosporins, and carbapenems. However, it primarily binds to penicillin-binding 3 of Gram-negative bacilli and does not bind to penicillin-binding proteins that are present in Gram-positive bacteria. Therefore, the spectrum of activity of aztreonam is limited to aerobic Gram-negative bacilli. Although this is very similar to the spectrum of aminoglycosides, these two classes of antibiotics are not interchangeable. Gram-negative bacteria can acquire resistance via the production of certain β-lactamases that hydrolyze aztreonam.

Aztreonam is not absorbed from the gastrointestinal tract. From a pharmacokinetic perspective, it can be given intramuscularly, but its frequency of administration (every 6–8 hours) makes this too painful to be practical, so only intravenous administration is useful. Therapeutic levels are achieved in nearly all body tissues and fluids,

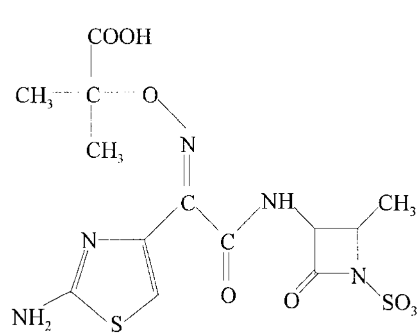


Figure 8–7 *Structure of aztreonam.*

including lungs and sputum, kidneys and urine, gallbladder, liver, bone, and prostate. Approximately 10% of aztreonam is hydrolyzed, presumably by hepatic enzymes, to yield metabolites that have the ring structure opened. Excretion is exclusively by renal mechanisms, via both glomerular filtration and tubular secretion.

The most common side effects are rashes. However, these are not due to immediate-type hypersensitivity. There is no cross-allergenicity between aztreonam and β-lactams and no risk of anaphylaxis to aztreonam in patients who have a history of anaphylaxis to those agents.

Glycopeptides

Despite being discovered 40 years ago, vancomycin has remained the only antibiotic in this class. It is a very large, complex glycopeptide structure with molecular weight of ~1450 daltons (the unit of measure). It acts by inhibiting bacterial cell wall synthesis. It surrounds the C-terminal D-alanyl-D-alanine precursor of the polypeptide portion peptidoglycan, thereby preventing hydrolysis of the last D-alanine of the polypeptide and the simultaneous transfer of the precursor to the growing cell wall by alanine transferase.

Until several years ago, acquired bacterial resistance to vancomycin did not occur. However, some strains of enterococci now produce an enzyme encoded by a gene called *van A*, which synthesizes a C-terminal D-alanyl-D-lactate precursor, to which vancomycin is unable to bind. When the terminal D-lactate is hydrolyzed, the resulting peptidoglycan is indistinguishable from that of vancomycin-susceptible strains.

Several cases of infections due to strains of *S. aureus* that are only intermediately susceptible to vancomycin have been reported. This may be only the beginning of a potentially huge clinical problem. The mechanism of decreased susceptibility by *S. aureus* has not yet been elucidated. More recently, several infections caused by *S. aureus* that are resistant to vancomycin have occurred. The mechanism of resistance is via the same *van A* gene as in enterococci. Prolonged previous treatment with vancomycin has occurred in all patients who have developed infections due to vancomycin-intermediate and vancomycin-resistant staphylococci.

Vancomycin is active against nearly all Gram-positive bacteria, including *Streptococcus* spp., *Enterococcus* spp., *Staphylococcus* spp., *Corynebacterium* spp., *Clostridium* spp., *Bacillus* spp., *Lactobacillus* spp., and *Listeria monocytogenes*. Gram-positive exceptions include *Leukonostoc* spp., some strains of enterococci, and rare strains of *S. aureus* (as previously noted) and *Staphylococcus haemolyticus*. It is inactive against Gram-negative bacteria. Therefore, it should be used when an infection due to methicillin-resistant *S. aureus* is suspected or proven, or when an immediate-type allergy to β-lactam antibiotics prevents the use of an antistaphylococcal penicillin.

Because it is a very large molecule, vancomycin is poorly absorbed. Intramuscular injections are too painful to be used for even a single dose. Therefore, except for the treatment of *C. difficile* colitis, it must be given intravenously. The usual dose of 1000 mg yields a peak serum concentration of 20 to 50 μg/mL and a trough of 5 to 12 μg/mL. Because of the relatively low toxic to therapeutic ratio, monitoring of trough levels should be performed to avoid side effects that are related to serum concentrations.

Adequate levels are achieved in skin, pleura, synovia, and other tissues that are susceptible to infection by Gram-positive bacteria. An exception to this general rule is the colonic lumen, which makes oral treatment necessary for *C. difficile* colitis. When being used to treat meningitis (e.g., due to methicillin-resistant *S. aureus*), additional intrathecal injections are sometimes necessary.

Vancomycin is excreted unchanged in urine via glomerular filtration, and therefore urinary levels are very high, but because it is not active against Gram-negative bacteria, they are usually irrelevant. The creatinine clearance of all patients who are to be given vancomycin should be calculated, because the dosage must be adjusted in patients with even mild renal insufficiency. In patients with renal failure who are on dialysis, it needs to be given only once every 2 to 7 days, depending on whether or not it is removed by the dialysis equipment being used (newer high-flux dialysis equipment removes large molecules, older machines do not).

If a dose is given too rapidly, flushing of the face and neck, sometimes accompanied by the "red man" or "red neck" syndrome, may occur. This is due to histamine release caused by local hyperosmolarity when a dose is given too rapidly, and is not an allergic reaction that precludes additional doses. Deafness may occur if serum levels are permitted to remain too high for a prolonged period. Older preparations of vancomycin were relatively impure and caused renal insufficiency. Many physicians mistakenly believe that vancomycin still can cause acute renal failure, but this is not a problem with the currently available preparations.

Aminoglycosides

Because of resistance by Gram-negative bacilli to many sulfonamides and early penicillins, which were the first two classes of antibiotics to be discovered, microbiologists began to screen soil bacteria for the production of other types of antibiotics. After several years of isolating antibiotics that were either too toxic or not sufficiently active to be useful, in 1943 a group headed by Seymour Waksman announced the discovery of an antibiotic that they named streptomycin, isolated from *Streptomyces griseus*. (All compounds that end in *-mycin* are derived from bacteria of the genus *Streptomyces*. This applies not only to aminoglycosides but also to other classes of antibiotics, such as erythromycin, and to many different types of medications, such as bleomycin, which is used to treat malignancies, and mithramycin, which is used to treat hypercalcemia.)

The aminoglycosides currently used in the United States are streptomycin, neomycin, kanamycin, gentamicin, tobramycin, and amikacin. Many species of bacteria that had been susceptible to streptomycin many years ago are now resistant, so it is now used mostly for the uncommon infections plague and tularemia, for synergy in serious enterococcal infections, and as second- or third-line agents for synergy in some cases of tuberculosis. Neomycin and kanamycin are only used topically and enterally because, compared with gentamicin, tobramycin, and amikacin, they are more toxic to humans when given systemically.

The chemical structure is a polycationic molecule consisting of at least two amino sugars linked by glycosidic bonds to an aminocyclitol ring (**Fig. 8–8**).

Aminoglycosides can be divided into chemical classes. In the first class, the aminocyclitol ring is streptidine. Streptomycin, which was the first aminoglycoside used, is the only member of this class. In the second chemical class, to which all the other aminoglycosides belong, the aminocyclitol ring is 2-deoxystreptamine. Within this second class are three families:

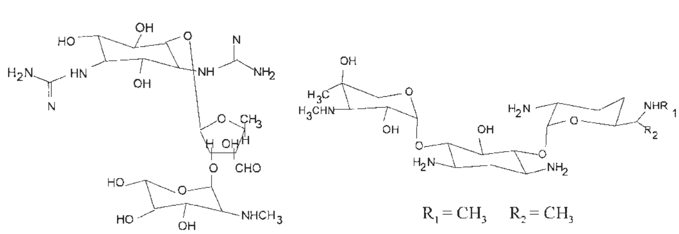
- Neomycin family
    - Neomycin: used orally for the treatment of hepatic encephalopathy (by killing bacteria that produce ammonium that is absorbed into the bloodstream) or topically on wounds
- Kanamycin family (one of the amino sugars is glucosamine)
    - Kanamycin: usually used orally to decontaminate the gastrointestinal tract prior to surgery
    - Tobramycin: slightly broader spectrum than gentamicin; used systemically
    - Amikacin: broadest spectrum of all aminoglycosides
- Gentamicin family (one of the amino sugars is garosamine)
    - Gentamicin: least broad spectrum of the aminoglycosides that are still commonly used, but still very active against many aerobic Gram-negative bacilli

Due to rapid, tight, and reversible binding to sites on the 30S subunit of the bacteria (for streptomycin) or both the 50S and 30S subunits (for the other aminoglycosides), aminoglycosides kill bacteria by inhibiting their protein synthesis and causing errors to be made in reading messenger ribonucleic acid (mRNA) codons. The binding appears to be in the region of the interface between the two ribosomal subunits.

Aminoglycosides must have at least one more site or mechanism of action, because they are bactericidal, in contrast to other antibiotics that inhibit protein synthesis, which are usually bacteriostatic. However, this additional site or mechanism has not been identified, and the exact nature of their interaction with bacteria remains incompletely understood.

Bacterial resistance is mediated by enzymes that are encoded on plasmids and mediate phosphorylation, acetylation, or adenylation of hydroxyl or amino groups on the amino sugars. This can either decrease aminoglycoside uptake by bacteria or decrease their binding to bacterial ribosomes, making them inactive. Several enzymes of each type have been found, each of which acts on a specific position and side group, and isoenzymes with different affinities for the same positions and side groups also exist. Other mutations encoding for changes in bacterial ribosomes or decreased uptake of aminoglycosides by bacteria may occur. These mechanisms of bacterial resistance are not as clinically important as are the aminoglycoside-modifying enzymes.

Aminoglycosides are active against facultatively anaerobic Gram-negative bacilli growing under aerobic conditions and aerobic Gram-negative bacilli. In combination with some penicillins or with vancomycin, streptomycin



STREPTOMYCIN GENTAMICIN C_1

Figure 8–8 *Structures of streptomycin and gentamicin C_1.*

and gentamicin are also used for synergy against *Enterococcus* spp. In combination with some penicillins and cephalosporins, they are also synergistic against Gram-negative bacilli. When used in this manner, the entry of streptomycin or gentamicin into bacteria is facilitated by the damage to the bacterial cell wall that is caused by the penicillin, cephalosporin, or vancomycin. Nowadays, aminoglycosides are almost always used in combination with another antibiotic (usually a penicillin, cephalosporin, carbapenem, or vancomycin) to treat serious infections due to Gram-negative bacilli or enterococci. Streptomycin, kanamycin, and amikacin are also active against *Mycobacterium tuberculosis.*

When given orally, absorption of aminoglycosides is very limited, and serum levels are too low to be practical to treat systemic infections. Nevertheless, kanamycin is given orally for decontamination of the gastrointestinal tract prior to gastrointestinal surgery, in which case absorption is not desired. Neomycin, which is too toxic to be given systemically, is used orally for hepatic encephalopathy, or topically on wounds, as mentioned earlier. When given intramuscularly, complete absorption of aminoglycosides occurs 30 to 90 minutes after injection.

Concentrations in bronchial secretions are only ~20% of those in serum, so aminoglycosides should not be used alone to treat patients with pneumonia. Levels in CSF, vitreous humor, bile, and prostatic tissue are also low. For the treatment of meningitis and ophthalmitis due to susceptible bacteria, intrathecal or intravitreal injections of aminoglycosides must be used, respectively.

Aminoglycosides are not metabolized by humans. Excretion occurs by incomplete filtration through glomeruli, with reabsorption of a small portion into renal proximal tubular cells. Eventually, nearly the entire dose is excreted in urine. Therefore, the creatinine clearance should always be calculated, the dose must be decreased for patients with even mild renal insufficiency, and serum concentrations should be monitored in all patients receiving aminoglycoside treatment. Peak levels correlate with efficacy, trough levels with toxicity. Less than 1% is excreted in bile.

With either a prolonged course of treatment or very high serum levels, reabsorption and pinocytosis into proximal tubular cells of kidneys results in high concentrations in these cells, causing a decrease in glomerular filtration rate that is manifested clinically by progressive, reversible, nonoliguric renal failure. Auditory or vestibular toxicity may also occur due to damage of the hair cells of the organ of Corti or crista ampullaris, respectively. This is thought to be related to the concentration of aminoglycoside in the endolymph or perilymph fluids of the inner ear. Although ototoxicity occurs much less frequently than nephrotoxicity, it is irreversible because it affects tissues that cannot regenerate. The different aminoglycosides have been shown to be primarily ototoxic, vestibulotoxic, or toxic to both systems (**Table 8–6**). Neomycin and kanamycin are examples of the aminoglycoside family that are primarily ototoxic. A very rare side effect specific to aminoglycosides is neuromuscular paralysis, which is due to both inhibition of presynaptic release of acetylcholine and blockade of postsynaptic receptors of acetylcholine. This is caused by high intrasynaptic concentrations of aminoglycoside and sometimes occurs when an intravenous dose is given very rapidly. If it is severe enough to interfere with respiration, it may be fatal.

Table 8–6 Auditory nerve toxity of aminoglycosides

Damage	Tobramycin	Gentamicin	Netilmicin
Cochlear	9.6%	16.4%	2%
Vestibular	3.3%	11.8%	0%

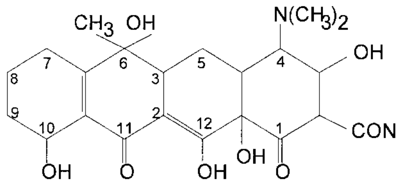
*Data from Lerner DA et al. Randomized controlled trial of the comparative efficacy, auditory toxicity, and nephrotoxicity of tobramycin and netilmicin. Lancet 1983;1:1123

Spectinomycin

Spectinomycin is chemically similar to aminoglycosides. Strictly speaking, it is an aminocyclitol (not an aminoglycoside) because it does not contain amino sugars or a glycosidic bond. It is the only member of the aminocyclitol group of antibiotics. Like aminoglycosides, it binds to the 30S subunit of the bacterial ribosome, but its mechanism of action is limited to inhibition of protein synthesis (i.e., it does not cause misreading of codons), and it is only bacteriostatic. Its action is limited to Gram-negative bacteria, but nowadays it is used only occasionally to treat gonorrhea because other, more effective antibiotics are available to treat infections due to other Gram-negative bacilli. Its pharmacokinetic characteristics are similar to those of aminoglycosides, but it does not have their associated nephrotoxicity and ototoxicity. It is given as a single intramuscular injection for uncomplicated gonorrhea when penicillinase-producing strains of *N. gonorrhoeae* (i.e., resistant to penicillin at a high level) are prevalent.

Tetracyclines

Screening for antibiotics produced by common soil bacteria led to the discovery of the tetracycline group of antibiotics in 1948. As expected, they had a different spectrum of activity than other classes of antibiotics because their chemical structure was different. The common types used today are tetracycline, oxytetracycline, doxycycline



Minocycline: H_2 on C6; $N(CH_3)_2$ on C7
Doxycycline: H & OH on C5; CH_3 & H on C6

Figure 8–9 *Structures of tetracycline, minocycline, and doxycycline.*

(Vibramycin), and minocycline (Minocin). For a brief discussion of demeclocycline (Declomycin), see the section on side effects.

As shown in **Fig. 8–9**, tetracyclines consist of four 6-carbon rings. The side chains that are on the rings determine properties such as rate of absorption, spectrum of activity, rate of excretion, and toxicity.

By binding to the 30S subunit of the bacterial ribosome, tetracyclines inhibit bacterial protein synthesis by preventing access of aminoacyl transfer RNA (tRNA) molecules to bacterial mRNA–ribosome–peptide complexes. Bacterial resistance is usually mediated by plasmids, which encode proteins that either interfere with active transport of tetracyclines into bacteria or cause active transport out of bacteria, keeping antibiotic levels inside bacterial cells very low.

Tetracyclines are useful in the treatment infections due to *Chlamydia* (e.g., urethritis, lymphogranuloma venereum, psittacosis, trachoma, inclusion conjunctivitis, pneumonia), *Mycoplasma, Rickettsia* (e.g., Rocky Mountain spotted fever, rickettsial pox, typhus), *Brucella, Vibrio* (including cholera), *Leptospira*, and *Borrelia* spp. (including Lyme disease in its early stages). Minocycline is also useful in the treatment of infections due to *Mycobacterium marinum* and *Nocardia* spp. Doxycycline has been used as an effective prophylaxis for traveler's diarrhea as well as for the treatment of acute epididymitis, nonspecific urethritis, and pelvic inflammatory disease.

Generally, the tetracyclines should not be used for the treatment of infections due to staphylococci, streptococci, or Gram-negative bacilli of the family Enterobacteriaceae because of the rapid emergence of resistance by these bacteria.

Following oral administration, tetracyclines are absorbed incompletely from the stomach and small intestine. The most important factor in determining absorption is the absence of the cations Ca^{+2}, Mg^{+2}, and Al^{+3}, which chelate tetracyclines and prevent their absorption. Therefore, patients must avoid eating dairy products and taking antacids (aluminum hydroxide) within about 2 hours of a dose of tetracycline. Tetracyclines should not be given intramuscularly because their injections are too painful.

Tetracyclines penetrate fairly well from blood into most body tissues. Even in the absence of meningeal inflammation, concentrations in CSF reach about one quarter of those in serum. High concentrations are also achieved in synovial fluid and in human milk. Excretion is primarily via kidneys, except for doxycycline, which is excreted in feces and therefore requires no dosage change even for patients with renal failure.

Most side effects of tetracyclines are reversible and include diarrhea, vaginal candidiasis, rashes, photosensitization, thrombophlebitis (when given intravenously), and worsening of renal failure that is already present. Two important side effects that may not be reversible should be noted: staining of teeth during their calcification (from a fetal age of about 5 months to about age 8 to 12 years in children), due to the binding of tetracyclines to calcium that was mentioned in the paragraph on their absorption; and fatty necrosis of the liver in pregnant women or in patients who receive 2000 mg or more per day. Demeclocycline can cause nephrogenic diabetes insipidus. It is no longer used as an antibiotic, but this side effect is exploited as a treatment for the syndrome of inappropriate antidiuretic hormone secretion (SIADH) when it does not respond to fluid restriction.

Chloramphenicol

There is only one member of this class of antibiotics, and it is also called chloramphenicol. The chemical structure is a relatively simple combination of nitrobenzene and a dichloroacetic acid derivative (**Fig. 8–10**). By binding to peptidyl transferase, a component of the 50S subunit of the bacterial ribosome, chloramphenicol prevents the binding of aminoacyl tRNA molecules to peptides that are being formed and inhibits bacterial protein synthesis. Bacterial resistance is almost always due to plasmid-encoded production of chloramphenicol acetyltransferase, which acetylates chloramphenicol molecules to

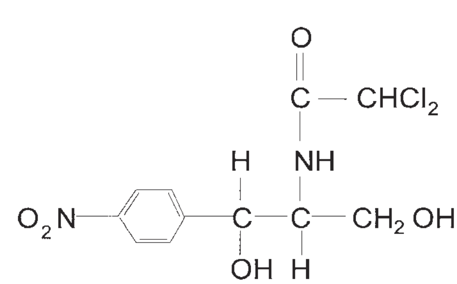


Figure 8–10 *Molecular structure of chloramphenicol.*

make them inactive. Occasionally, resistance results due to the loss of antibiotic permeability into bacteria.

Chloramphenicol has a very wide spectrum of activity, including most aerobic Gram-positive bacteria (except *Enterococcus* spp. and methicillin-resistant *S. aureus*), most aerobic Gram-negative bacteria (except *Pseudomonas* spp., *Acinetobacter* spp., and a few others), almost all anaerobic bacteria, and *Rickettsia* spp.

The palmitate formulation, which is used for oral dosing, is very well absorbed after oral administration, and a peak serum concentration of ~12 μg/mL is achieved in 1 hour. Intramuscular administration results in slower absorption and lower peak serum concentrations than intravenous dosing and should be avoided for these reasons. The succinate formulation, which is used for intravenous administration, is hydrolyzed in the liver to yield free chloramphenicol, but because the hydrolysis is not complete, peak concentrations are only 70 to 80% of those obtained with oral doses. Therefore, oral administration is the preferred route whenever possible, even for severe infections. Because it is very lipophilic, chloramphenicol is well distributed in almost all body tissues, including the CNS.

Chloramphenicol is primarily glucuronidated to an inactive metabolite by the liver, making it more soluble in water so that it can be excreted in urine. Because of their impaired ability to glucuronidate, chloramphenicol should be used with caution to treat infections in newborn infants. Following glucuronidation, chloramphenicol is secreted into urine. Small amounts of unmetabolized (active) chloramphenicol are also filtered into urine and can treat urinary tract infections.

Bone marrow aplasia, which occurs in ~1 in 20,000 to 40,000 individuals, is the most feared side effect. It usually occurs several weeks to months following a course of treatment with chloramphenicol, and it is not possible to predict in whom it will occur. It is usually irreversible and fatal, and can occur with any route of administration. Because of this and the availability of other classes of effective antibiotics for the past 2 decades, chloramphenicol is only rarely used nowadays despite its very broad spectrum of activity. A more common effect on bone marrow is a mild dose-related, reversible depression of function, which resolves after discontinuation of the antibiotic. Gray baby syndrome is another side effect and is due to the inability of neonates to glucuronidate chloramphenicol and the consequent failure to excrete the unconjugated antibiotic. It is characterized by hypotension, cyanosis, and abdominal distension, and is associated with extremely high levels of chloramphenicol in the bloodstream. In newborns, serum levels of chloramphenicol should be monitored to prevent this side effect. Other infrequent side effects include optic neuritis, gastrointestinal intolerance, fever, and rash.

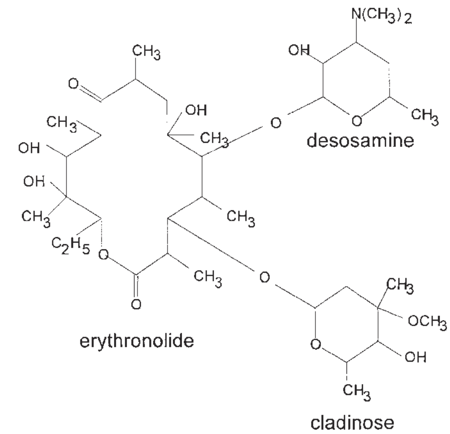


Figure 8–11 *Structure of erythromycin.*

Erythromycins (Macrolides)

Erythromycin has been in use as an antibiotic for many years. The newer members of the erythromycin family, azithromycin (Zithromax) and clarithromycin (Biaxin), have been available since 1991 and are used extensively. The chemical structure of the erythromycins is a many-membered lactone ring to which sugars are attached (**Fig. 8–11**).

In the absence of aminoacyl tRNA, erythromycins reversibly bind to the 50S subunit of the bacterial ribosome, thus preventing synthesis of new proteins. Their binding site may be the same as that of chloramphenicol (i.e., peptidyl transferase) because they interfere with the binding of chloramphenicol to bacterial ribosomes. Bacterial resistance is via decreased permeability of the cell envelope (in Enterobacteriaceae), plasmid-mediated methylation of adenine residues in the 23S RNA component of the 50S subunit of the bacterial ribosome, and a chromosomal mutation affecting a protein in the 50S subunit (possibly peptidyl transferase).

Erythromycins are effective antibiotics for the treatment of pneumonia due to *Mycoplasma* spp. and *Legionella* spp. In patients who are allergic to penicillin, it is an alternative antibiotic in treating group A β-hemolytic streptococci (*S. pyogenes*), pneumococci (*S. pneumoniae)*, chlamydial infections, cutaneous staphylococcal infections, syphilis, and gonorrhea. When given to patients early in the course of either gastroenteritis due to *Campylobacter jejuni* or whooping cough (*Bordetella pertussis*), it may shorten the duration of the illness.

Clarithromycin and azithromycin are active against several additional organisms, including *H. influenzae,* making them more useful than the old formulations of erythromycin in the treatment of some cases of otitis, sinusitis, and atypical pneumonia. They are also both active against *Mycobacterium avium-intracellulare* and *Toxoplasma gondii*. Azithromycin can be used to treat urethritis due to *Chlamydia trachomatis*.

Oral formulations of erythromycin are moderately absorbed from the gastrointestinal tract. The peak serum concentration after a dose of 500 mg is 1 to 2 μg/mL. Absorption of the estolate formulation is increased by food in the stomach; the opposite is true for the base, stearate, and ethylsuccinate forms. Intramuscular injections are too painful to be used. The lactobionate form of erythromycin is used for intravenous injection. Oral clarithromycin and azithromycin are also absorbed well. Since early 1998, an intravenous formulation of azithromycin has also been available, increasing its use for patients requiring hospitalization for pneumonia.

Erythromycins diffuse well into most body tissues, with the exception of the CNS. Erythromycin crosses the placenta and is also present in human milk. Most of an administered dose is metabolized by the liver to a demethyl form, or is metabolized in tissues. On a per weight basis, azithromycin is less active than either erythromycin or clarithromycin, but its extremely high levels in tissues (except for the CNS) make it just as effective against susceptible organisms. The metabolism of azithromycin is very slow, so it should be given only once a day. Only a small fraction (5–15% for erythromycin) is excreted in urine or feces. The dose does not have to be decreased for patients with kidney failure.

Erythromycins are relatively nontoxic. The major side effects are abdominal cramps and diarrhea due to an increase in intestinal motility and local thrombophlebitis when given intravenously, but these are not serious. Cholestatic hepatitis is occasionally caused only by the erythromycin estolate formulation. Deafness can occur with high doses, but this is not due to nerve damage, and hearing returns to normal after the dose is stopped.

Ketolides

The first ketolide, telithromycin (Ketek), was approved for use in 2004. The ketolide group is considered a new class of antibiotics, although similar to the erythromycins, with a ketone group replacing one of the sugars (cladinose). The cladinose in erythromycins is responsible for inducing an enzyme that causes resistance. In the United States, 20 to 30% of *S. pneumoniae* strains currently are resistant to erythromycins. Telithromycin not only binds to the same site on bacterial ribosomes as do erythromycins but has one additional binding site and is therefore significantly more active than macrolides against *S. pneumoniae* (including strains that are resistant to other classes of antibiotics). It also has activity against *H. influenzae*, *M. catarrhalis*, *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, and *Legionella pneumophila*. It is available in tablet form, given once a day, and its bioavailability is 57% of a total dose. Levels well above those necessary to inhibit more than 90% of *S. pneumoniae* and *H. influenzae* strains have been measured in both sinus tissue and bronchial epithelial lining fluid. Efficacy has been demonstrated in acute sinusitis, mild to moderate community-acquired pneumonia, and exacerbations of chronic bronchitis. Side effects are generally similar to those of the macrolides, although it can occasionally also interfere with release of visual accommodation (i.e., when looking from near to far). Interactions with other medications are theoretically the same as those seen with clarithromycin due to their similar inhibition of the cytochrome P-450 enzyme system.

Lincosamides

Clindamycin (Clecocin) is the only member of this class of antibiotics. It is the result of the addition of a chlorine atom to lincomycin, which was the first lincosamide and is no longer used in the United States. Its chemical structure consists of two sugars linked by an amide bond (**Fig. 8–12**).

The mechanism of action of clindamycin is inhibition of bacterial protein synthesis via binding to peptidyl transferase in the 50S subunit of the bacterial ribosome (possibly the same site as chloramphenicol and the erythromycins). Bacterial resistance is both via chromosomally mediated alteration of a protein in the 50S subunit of the bacterial ribosome and plasmid-mediated methylation of adenine residues in 23S ribosomal RNA of the 50S ribosomal subunit. These are the same mechanisms of bacterial resistance to erythromycins, and cross-resistance between lincosamides and erythromycins

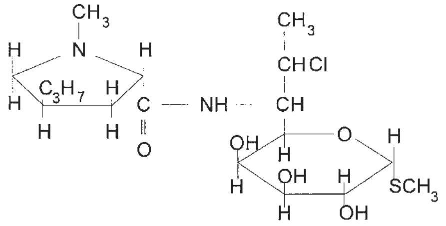


Figure 8–12 *Structure of clindamycin.*

is common. Additionally, in *S. aureus*, inactivation by plasmid-mediated adenylation of lincosamides occurs. Its spectrum of activity is anaerobic bacteria (except *C. difficile*), Gram-positive aerobic bacteria (except *Enterococcus* spp. and methicillin-resistant *Staphylococcus aureus*), and some protozoa (*Toxoplasma gondii, Plasmodium* spp., *and Babesia microti*). The anaerobic and Gram-positive portions of its antibacterial spectrum make it useful in treating cases of chronic sinusitis and otitis, in which these bacteria are more frequently encountered.

Clindamycin is moderately well absorbed after oral administration, with a peak serum concentration of 3.6 μg/mL after a 300 mg dose. Serum levels of 11 to 14 μg/mL are achieved after intramuscular or intravenous administration of 900 to 1200 mg. The only tissue into which clindamycin does not appreciably penetrate is the CNS. Most of an administered dose is metabolized in the liver to N-demethyl clindamycin and clindamycin sulfoxide.

Only ~10% of a dose is eliminated in urine, and a smaller amount is eliminated in feces. The metabolites can be found in both urine and feces as well. Because of the dual elimination, the dosage has to be reduced only in cases of severe renal failure.

The important adverse effects of clindamycin are toxin-mediated colitis caused by *C. difficile*, and diarrhea that is probably due to an alteration of colonic flora. The colitis may be severe, even fatal. It can be diagnosed by an assay for the toxin in a stool specimen or by visualizing pseudomembranes on colonic mucosa during colonoscopy. It usually improves simply by stopping the clindamycin, but if this is not possible, it can be treated with either oral metronidazole or vancomycin. *C. difficile* colitis can occur with the administration of any antibiotic, and nowadays it actually occurs more often due to other types of antibiotics, but it was first recognized with clindamycin.

Metronidazole

Metronidazole (Flagyl) is the only member of the nitroimidazole group of antibiotics. It has a simple chemical structure that is shown in **Fig. 8–13**. Its mechanism of action is via reduction of the nitro group on the

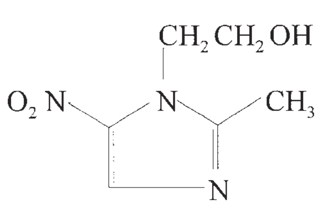


Figure 8–13 Structure of metronidazole.

molecule, which generates short-lived free radicals that damage cellular structures. Acquired bacterial resistance to metronidazole is rare and may result from both decreased uptake of the molecule and decreased reduction of the nitro group. Originally, it was developed for the treatment of infections due to *Trichomonas vaginalis*, but it was also recognized as being active against infections caused by anaerobic bacteria, including *Bacteroides* spp., *Prevotella* spp., *Fusobacterium* spp., *Clostridium* spp., and anaerobic streptococci.

After oral administration, it is absorbed very quickly and nearly completely, with a peak serum level of ~25 μg/mL, which is comparable to that obtained with intravenous dosing and easily above the level needed to inhibit susceptible bacteria. Therapeutic concentrations are achieved in all body fluids and tissues. Hydroxylation (which produces an active metabolite), acetylation, glucuronidation, and glucuronidation of the hydroxyl derivative are the pathways by which humans metabolize metronidazole, with 60 to 80% of a dose excreted unchanged and as metabolites in urine and the remainder in feces. The dose needs to be decreased only in cases of severe combined renal and hepatic failure.

Serious side effects of metronidazole are rare. Nausea, anorexia, and other gastrointestinal symptoms occasionally occur, as can a metallic taste. It can also cause a disulfiram-like reaction in people who drink alcohol, so patients should be instructed not to drink liquor while taking a course of metronidazole. It has been known for many years that metronidazole is mutagenic in the Ames test and has induced tumors in mice that have received it for long periods of time, but retrospective and prospective studies of humans have not demonstrated a link to neoplasms.

Oxazolidinones

Linezolid (Zyvox) is the only member of this class and has been on the market since 2000. It is a relatively small molecule with both hydrophilic and hydrophobic portions and inhibits bacterial protein synthesis by binding to the 50S subunit of the bacterial ribosome. Bacterial resistance occurs as a result of altered ribosomal RNA in the 50S ribosomal subunit, which results in decreased binding by linezolid.

Its spectrum of activity is strictly Gram-positive cocci, including all streptococci, and nearly all staphylococci (including methicillin-resistant strains) and enterococci (including vancomycin-resistant strains). It is inactive against Gram-negative rods and strict anaerobes.

Its chemical characteristics allow for oral administration with 100% absorption, although an intravenous formulation is also available for patients whose gastrointestinal tracts are nonfunctional. Therapeutic concentrations are achieved in most body fluids and tissues, including skin, skin structures, lungs, bronchoalveolar fluid, and saliva. Metabolism is via oxidation and occurs extensively in the liver, with ~50% of a dose then being excreted metabolized in urine, 30% unmetabolized in urine, 10% metabolized in feces, and 10% undetermined.

The most common side effects are mild gastrointestinal disturbances such as nausea or vomiting. In clinical trials, thrombocytopenia occurred more often with linezolid than with comparator antibiotics, but the mechanism is not clear. Linezolid is also a reversible and nonselective monoamine oxidase inhibitor (MAOI), so there is potential for interactions with adrenergic agents (e.g., pseudoephedrine) and tyramine-containing foods (aged, fermented, pickled, or smoked foods such as certain cheeses, meats, wines, and beers); however, these have not been observed clinically.

Despite its excellent activity against Gram-positive cocci, linezolid should not be used for routine infections caused by these bacteria unless resistance to older antibiotics is suspected; otherwise, bacterial resistance to linezolid as well will become a significant problem relatively quickly. Resistance to linezolid by staphylococci and enterococci has already been observed despite the relatively short time that it has been available.

Quinolones

The quinolones are a class of antibiotic that has been available since 1986. They are chemically related to nalidixic acid (a naphthyridine antibiotic first used in 1962) and possess several important properties: after oral administration, they achieve serum and tissue concentrations that are above their minimum inhibitory concentrations for many nosocomial strains of bacteria; they are active against many bacteria that are resistant to β-lactams and aminoglycosides; and they have relatively little toxicity. The most frequently used types are ciprofloxacin (Cipro), levofloxacin (Levaquin), and gatifloxacin (Tequin). Levofloxacin is the levorotary isomer of ofloxacin (Floxin), which is also available for clinical use.

All quinolones have an important carboxyl group (and are sometimes called carboxyquinolones) and a fluorine atom (they are also called fluoroquinolones), and most also have a piperazine ring. The fluorine and piperazine significantly increase the antibacterial activity of quinolones, and the carboxyl group is necessary for activity.

The antibacterial activity of quinolones is due to the inhibition of deoxyribonucleic acid (DNA) gyrase, the enzyme that puts supercoils into circular bacterial DNA at a site distant from the replication fork during DNA replication so that the DNA can unwind at the replication fork as it is being duplicated.

Although quinolones enter bacteria by passive diffusion, they are pumped out by an active system that requires energy. Bacterial resistance to quinolones occurs by both chromosomal mutations that cause an increase in the rate of active efflux from bacteria and mutations that cause alterations in DNA gyrase so that quinolones cannot bind to it. The frequency of resistance (mutation) is dependent on the species of bacteria; for instance, it occurs more commonly in *P. aeruginosa* and *S. aureus* than in *E. coli* or other species in the Enterobacteriaceae family.

Quinolones are very active against Gram-negative bacilli, including Enterobacteriaceae and *P. aeruginosa*, and are also very active against Gram-negative cocci such as *H. influenzae*, *Neisseria* spp., and *M. catarrhalis*. Levofloxacin, gatifloxacin, and moxifloxacin (Avelox) are the most active of this group against Gram-positive cocci (staphylococci and streptococci); generally, the other quinolones should not be used to treat infections caused by these bacteria. Ciprofloxacin has the best activity among quinolones against *P. aeruginosa*. Levofloxacin, gatifloxacin, and moxifloxacin are active against *L. pneumophila, M. pneumoniae,* and *C. trachomatis*, making them useful for the empirical treatment of pneumonia.

Quinolones are rapidly and nearly completely absorbed from the gastrointestinal tract. The bioavailability is 70% for oral ciprofloxacin, 99% for oral levofloxacin, and 95% for oral gatifloxacin. The availability of oral formulations with good absorption combined with a very broad spectrum of activity has created a large potential for overuse in the outpatient setting, with consequent development of bacterial resistance.

Absorption of some quinolones is reduced by antacids, iron, zinc-containing vitamins, and sucralfate. After absorption, quinolones are widely distributed into tissue and fluids. Data regarding efficacy of quinolones in treating infections of the CNS are very limited to date, and other antibiotics should be used to treat meningitis. Therapeutic concentrations against Gram-positive cocci may not be achieved in CSF.

The type of hepatic metabolism of active quinolones is variable, depending on the particular quinolone, and includes glucuronidation, carboxylation, hydroxylation, and demethylation. Variable amounts are recovered either unmetabolized in urine and feces (up to 60% and 10%, respectively, for ciprofloxacin) and as metabolites in urine and feces.

The most common side effects of quinolones are nausea, vomiting, abdominal pain, headache, dizziness, seizures, and rashes. Most of these side effects are mild, and their overall incidence is low. All quinolines also rarely cause prolongation of the QT interval of the electrocardiogram, which may lead to torsades de pointes and ventricular fibrillation, and may cause tendon rupture, usually around a weight-bearing joint such as the Achilles tendon. In animal testing, early quinolones caused cartilage damage to beagle puppies, and therefore this class of antibiotics should be used in children only when there are no alternatives.

Antifungals

With the advent of both natural and artificial forms of immunodeficiency in the last 2 decades, an increasing percentage of the population is becoming more susceptible to a variety of fungal infections. This includes patients with acquired immunodeficiency syndrome (AIDS) as well as those who receive myelosuppressive chemotherapy and invasive catheters for a multitude of malignancies and other underlying conditions that require relatively new "life-saving" treatment and monitoring. Although most fungal infections are treated by internists and specialists in infectious diseases, and most fungi that grow in cultures sent by otorhinolaryngologists are colonizers, otorhinolaryngologists should nevertheless possess a basic understanding of frequently used antifungal agents because of occasional patients under their care with conditions such as oropharyngeal candidiasis, invasive sinusitis due to *Aspergillus* spp., or rhinocerebral mucormycosis.

Polyenes

This group is constituted by two agents, amphotericin B and nystatin. These are extremely large molecules composed of multiple units of isoprene and a ring structure. Both act by binding to sterols, most notably ergosterol, in fungal cytoplasmic membranes. Acquired fungal resistance to polyenes is rare and is the result of alterations in fungal sterols. Because these are essential structural components, fungi with acquired resistance to polyenes may not be as pathogenic as those that are susceptible.

Polyenes have a very broad spectrum of antifungal activity. Nearly all fungi are susceptible to amphotericin B, which has made it the mainstay of treatment for serious fungal infections for the past 4 decades. Due to the large size of their molecules, neither amphotericin B nor nystatin is absorbed when given orally. Both are also very hydrophobic, and therefore insoluble in water. The intravenous formulation of amphotericin B deoxycholate (Fungizone) that has been on the market for many years is a colloid, not a solution. Three relatively new formulations of amphotericin B for intravenous administration are also on the market: amphotericin B colloidal dispersion (ABCD, amphotericin B in disk-like structures, Amphotec), amphotericin B lipid complex (ABLC, amphotericin B in ribbon-like structures, Abelcet), and liposomal amphotericin B (amphotericin B in unilamellar liposomes, Ambisome). Nystatin (Mycostatin, Nilstat) is available only in oral and topical formulations.

The steady-state half-life of amphotericin B is ~7 days, and dosing does not have to be changed for patients in renal failure. Penetration into CSF is poor, although measurable concentrations can be found in meningeal tissue itself. The disposition of polyenes following intravenous administration is unclear. Most of a dose is not recoverable in either urine or feces, either unchanged or as metabolites. It is probably extensively incorporated into human cell membranes via binding to sterols, including cholesterol.

The major problem with amphotericin B is not a shortfall in efficacy, but its high toxicity. As a result, it is indicated only for severe, life-threatening fungal infections. Even its once-daily frequency of administration is not dictated by its half-life, as with most medications, but by its side effects. These include fevers and chills, dose-dependent glomerular damage and renal insufficiency that are sometimes irreversible, potassium wasting, phlebitis at the infusion site, and anemia. (Intravenous nystatin never made it to the market because its side effects were even more severe than those caused by amphotericin B, even in lipid-based formulations.) The newer lipid-based formulations of amphotericin B cause renal insufficiency significantly less frequently than amphotericin B deoxycholate while retaining comparable efficacy. However, they still cause fever and rigors at very high rates, and their enormous cost has precluded their displacing amphotericin B deoxycholate.

Azoles

This class of antifungal agents has enabled physicians to treat some serious fungal diseases without the side effects of amphotericin B, frequently with oral therapy. Fluconazole (Diflucan), itraconazole (Sporanox), and voriconazole (Vfend) are the most commonly used types. They act by inhibiting the fungal cytochrome P-450 enzyme 14-α demethylase, which is responsible for the conversion of lanosterol to ergosterol. Reduced concentrations of ergosterol and elevated levels of 14-α sterols including lanosterol can be measured in fungal cell membranes. Alteration of the target demethylase enzyme can result in acquired fungal resistance.

All three commonly used azoles are active against most *Candida* spp., although in the case of fluconazole

C. krusei is notable for being highly resistant and *C. glabrata* often has partial resistance. Fluconazole is also indicated for the treatment of cryptococcal meningitis, where it is usually used as maintenance treatment after amphotericin B has been used. Itraconazole is also indicated for the treatment of blastomycosis and histoplasmosis. *Aspergillus* spp. are intrinsically resistant to fluconazole, but patients with aspergillosis sometimes clinically respond to itraconazole, and voriconazole was found to be superior to amphotericin B in one study of patients with invasive aspergillosis. Some molds, for example, *Rhizopus* spp. and other causes of mucormycosis, are inherently resistant to all azoles.

The bioavailability of oral fluconazole is ~90%, and that of itraconazole tablets is 55% following a full meal and only 20% on an empty stomach. An oral suspension of itraconazole is also available; this preparation yields a bioavailability of ~80% when taken on an empty stomach. Oral voriconazole has a bioavailability of 96%.

Azoles are distributed to most body tissues and fluids, with the exception of the CSF in the case of itraconazole. Approximately 80% of a fluconazole dose is excreted unchanged in urine, and 10% is excreted as a variety of metabolites. The dose must be decreased for patients with renal insufficiency. In the case of itraconazole, most of a dose is metabolized, with less than 1% excreted unchanged in urine, 40% excreted as a variety of metabolites in urine, and 5 to 10% excreted as unchanged metabolites in stool; the dose does not require an adjustment for patients with renal failure.

Most of the side effects of itraconazole and voriconazole are due to interactions with other medications that are metabolized by the human cytochrome P-450 system. These interactions do not occur with fluconazole. Terfenadine (Seldane), astemizole (Hismanal), cisapride (Propulsid), oral hypoglycemics, warfarin, phenytoin (Dilantin), cyclosporin (Sandimmune, Neoral), rifampin, and theophylline all should not be given with azoles, or, if they must be given, their levels and clinical effects should always be carefully monitored.

Suggested Readings

Brody TM, Larner J, Minneman KP, eds. Human Pharmacology Molecular to Clinical. 3rd ed. St. Louis: CW Mosby; 1998

Carpenter RL, Mackey DC. Local anesthetics. In: Barash PG, Cullen BF, Stoelting RK, eds. Clinical Anesthesia. 3rd ed. Philadelphia: Lippincott-Raven; 1997:413–440

Catterall WA, Makie K. Local anesthetics. In: Hardman JG, Limbind LE, Gilman AG, eds. The Pharmacological Basis of Therapeutics. 10th ed. New York: McGraw-Hill; 2001:367–384

Chatting BG. Basic and Clinical Pharmacology. 5th ed. Norwalk, CT: Appleton and Lang; 1992

Clark WG, Brater DC, Johnson AR, eds. Goth's Medical Pharmacology. 12th ed. St. Louis: CW Mosby; 1988

Coda BA. Opioids. In: Barash PG, Cullen BF, Stoelting RK, eds. Clinical Anesthesia. 3rd ed. Philadelphia: Lippincott-Raven; 1997:329–358

Gorbach SL, Bartlett JG, Blacklow NR. Infectious Diseases. 3rd ed. Philadelphia: WB Saunders; 2004

Gotta AW, Donovan R, Sullivan CA. The pharmacology of local anesthetics. Ophthalmol Clin NA 1998;11(1):11–23

Gutstein HB, Akil H. Opioid analgesics. In: Hardman JG, Limbind LE, Gilman AG, eds. The Pharmacological Basis of Therapeutics. 10th ed. New York: McGraw-Hill; 2001:569–620

Hardman JG, Limbind LE, Gilman AG, eds. The Pharmacological Basis of Therapeutics. 10th ed. New York: McGraw-Hill; 2001:ch. 1–3

Kaufman JA. The otolaryngologic manifestations of gastroesophageal reflux disease (GERD). Laryngoscope 1991;101 (4, suppl 53)

Kucers A, Crowe S, Grayson ML, Hoy J. The Use of Antibiotics. 5th ed. London: Butterworth-Heinemann; 1997

Mandell GL, Bennett JE, Dolin R. Principles and Practice of Infectious Diseases. 6th ed. London: Churchill Livingstone; 2004

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. Biotransformation of drugs take place predominantly in the
   - A. Smooth endoplasmic reticulum of the liver
   - B. Renal tubules
   - C. Lung alveoli
2. Drugs pass more easily across the placenta if they are
   - A. Lipophilic
   - B. Polar
3. The antacid that has been shown to interfere with the cytochrome P-450 system is
   - A. Ranitidine
   - B. Cimetidine
   - C. Omeprazole
4. All of the following are amide local anesthetics except
   - A. Lidocaine
   - B. Bupivacaine
   - C. Ropivacaine
   - D. Tetracaine
   - E. Mepivacaine

5. A unique characteristic of bupivacaine-induced toxicity is
   A. Intense drowsiness
   B. Central nervous system toxicity often precedes cardiac toxicity.
   C. Ventricular tachycardia
   D. Sudden onset of congestive heart failure
   E. Atrial fibrillation
6. Ester local anesthetics are metabolized by the same enzyme that metabolizes
   A. Curare
   B. Morphine
   C. Succinylcholine
   D. Meperidine
   E. Lidocaine
7. The analgesic effect of codeine is due to its metabolism to
   A. Meperidine
   B. Morphine
   C. Normeperidine
   D. Heroin
   E. Fentanyl
8. Meperidine is contraindicated in patients with renal failure because its metabolite, meperidine, can
   A. Cause convulsions
   B. Precipitate congestive heart failure
   C. Cause ventricular arrhythmias
   D. Cause anaphylaxis
   E. Precipitate narcotic withdrawal
9. The analgesic and respiratory depressant effects of the opioids can be reversed with
   A. Naloxone
   B. Normeperidine
   C. Oxymorphone
   D. Morphine glucuronide
   E. Fentanyl
10. Which of the following antibiotics is most likely to cause diarrhea?
    A. Metronidazole
    B. Clindamycin
    C. Gentamicin
    D. Tetracycline
    E. Imipenem
11. Which of the following antibiotics is best suited for empirical use in acute uncomplicated sinusitis?
    A. Amoxicillin-clavulanic acid
    B. Tetracycline
    C. Tobramycin
    D. Ertapenem
    E. Cephradine (first-generation cephalosporin)
12. When treating a hospitalized, seriously ill patient with an ear, nose, or throat infection that requires empirical coverage for streptococci, *Haemophilus influenzae*, and nosocomial enteric Gram-negative bacilli, which of the following antibiotics is best to use?
    A. Penicillin V
    B. Amoxicillin-clavulanic acid
    C. Metronidazole
    D. Doxycycline
    E. Ceftriaxone

Chapter 9
OTOTOXICITY

LEONARD P. RYBAK, JOHN S. TOULIATOS, AND KATHLEEN CAMPBELL

Certain drugs have the capacity to damage the inner ear, causing auditory or vestibular dysfunction. A basic understanding of the drugs known to have this propensity, as well as an understanding of the means available to detect ototoxicity in a patient, is essential to the physician prescribing ototoxic drugs and managing the patient with ototoxicity. This knowledge can alert the astute clinician that specific inner ear dysfunction is a direct consequence of ototoxic drug therapy, potentially altering therapy to prevent further damage. Monitoring of patients who require treatment with known ototoxic drugs can prevent or reduce inner ear toxicity as well as allay patients' anxiety regarding the possible side effects. This knowledge of ototoxicity has obvious medicolegal consequences and will influence decisions regarding cochlear implants. Finally, understanding the physiology underlying the toxicity of certain drugs can lead to future development of otoprotectant agents and elimination of ototoxic drugs.

FREE RADICALS

Free radicals are reactive chemical compounds. A free radical is any atom or molecule containing one or more unpaired electrons that have escaped the solvent cage or enzyme active site in which they were generated. They can be detected directly or indirectly by chemical and physical methods. Antioxidants are chemicals that protect against free radicals by trapping the radical so that it is no longer free to cause tissue damage. Reactive oxygen species or free radicals may be generated by several processes that are damaging to the inner ear. Ischemia and reperfusion injury, acoustic trauma, irradiation therapy, and ototoxic injury may be remediated, at least in part, by the production of reactive oxygen species in the inner ear. These intermediates may then react with membrane lipids, proteins, and other cellular components to cause damage and death to hair cells and other cells in the inner ear. The knowledge that this may occur suggests the use of protective agents that can spare the antioxidant enzymes and glutalhione in the inner ear. Some specific therapeutic interventions are mentioned in text under specific ototoxic agents.

This chapter discusses current knowledge concerning ototoxic medications, monitoring of ototoxic drugs, and management of patients with ototoxicity.

OTOTOXIC MEDICATIONS

Aminoglycosides

Aminoglycosides were the first chemical substances to draw widespread attention to the problem of drug-induced hearing loss after being used liberally in the 1940s to combat tuberculosis. The incidence of clinical hearing loss secondary to aminoglycoside therapy has been diminished significantly because of the use of newer derivatives with lower ototoxic potential, the advent of potentially less toxic alternatives (third-generation cephalosporins/quinolones), and efficient monitoring of serum levels. Regardless, this problem continues to exist and is exacerbated in many other countries, with estimates that aminoglycoside-induced hearing loss is responsible for up to 66% of deaf-mutism in China.

Aminoglycosides have bactericidal activity primarily against gram-negative aerobic bacilli. They act on bacterial ribosomes and stop the synthesis of bacterial-cell protein, causing accumulation of intermediate metabolic products that are toxic to the bacterial cell. The antibiotic is cleared through the kidneys; therefore, dosing schedules must be adjusted in patients who have decreased renal function. The more commonly known agents include streptomycin, kanamycin, gentamicin, netilmicin, tobramycin, amikacin, and neomycin. With all these drugs, both auditory and vestibular toxicity can occur simultaneously. However, streptomycin and gentamicin are primarily vestibulotoxic, whereas amikacin, neomycin, and kanamycin are primarily cochleotoxic. A correlation between chemical structure and preferred site of toxic effect has not yet been established. The selective vestibulotoxicity effects of aminoglycosides have been exploited both parenterally and transtympanically to treat select cases of disabling vertigo caused by Meniere's disease. Neomycin and kanamycin are restricted to topical application due to severe ototoxic and nephrotoxic effects with parenteral use. Although poorly absorbed in the gastrointestinal (GI) tract because they are highly polar cations, there have been reports of ototoxicity with oral administration, and toxicity is heightened in patients with GI disease.

Aminoglycoside-induced hearing loss begins at the high frequencies and is related to progressive and irreversible destruction of outer hair cells of the cochlea. Hair cell destruction begins at the basal coil and proceeds to the apex, and although the efferent cells are first affected, the inner hair cells eventually are damaged with progressive toxicity. Vestibulotoxicity can produce oscillopsia and disequilibrium related to damage in the vestibular sensory neuroepithelium, with most extensive hair cell damage beginning in the apex of the ampullar cristae and the striolar regions of the maculae of the saccule and utricle. The underlying cellular mechanisms responsible for the irreversible hair cell loss remain uncertain, although several theories exist, and plentiful research into the mechanisms of action of these drugs is ongoing.

Acute and chronic toxicity appear to have distinct underlying mechanisms. In acute reversible toxicity, it appears that calcium antagonism may play a role, with subsequent blockage of ion channels in efferent cells. At the present time, the most promising mechanism for chronic aminoglycoside ototoxicity appears to be the following: the aminoglycoside forms a complex with iron, and this complex produces free radicals. The free radicals can interact with membranes and proteins of cells in the cochlea and produce irreversible damage. Recently, a genetic predisposition to aminoglycoside-induced ototoxic deafness has been linked to a mitochondrial mutation that can be inherited through the maternal lineage.

As the cellular mechanisms underlying aminoglycoside ototoxicity continue to be unveiled, concurrent studies testing potential otoprotective agents such as antioxidants and iron chelators are ongoing, with hopes to eradicate ototoxic effects of aminoglycosides in the future.

Loop Diuretics

The loop diuretics are a group of potent synthetic drugs that act on the Henle's loop to inhibit the reabsorption of sodium, potassium, and chloride ions. The drugs are used to treat congestive heart failure, hypertension, edema of renal failure, cirrhotic ascites, and certain pulmonary disorders in the newborn. Potentially reversible yet sometimes permanent ototoxic effects can develop with their use.

Loop diuretics undergo renal excretion, and plasma half-life not only is prolonged in renal failure but also has been documented with congestive heart failure. Plasma levels exceeding 50 mg/L frequently are associated with middle- and high-frequency sensorineural hearing loss, tinnitus, and vertigo. The two major ototoxic diuretics are furosemide and ethacrynic acid. Increased ototoxic effects have been reported with concomitant aminoglycoside treatment.

Drug-induced changes in the electrolyte composition of the endolymphatic fluid occurs with this class of drugs. Damage to the stria vascularis and the dark cell areas of the vestibular system are the most pronounced histological changes documented with loop diuretic ototoxicity. Outer hair cell loss in the basal turn of the cochlea has been demonstrated in some studies.

Antineoplastic Agents

Cisplatin

Of the chemotherapeutic agents that have been reported to affect the inner ear adversely, cisplatin is the major ototoxic agent among these compounds. Cisplatin is a potent antineoplastic agent that has been used widely for a variety of malignancies since its approval by the U.S. Food and Drug Administration in 1978. It works by inserting into the helix of deoxyribonucleic acid (DNA) and halting replication. In addition to ototoxicity, peripheral neuropathy and renal insufficiency can develop with cisplatin therapy. Ototoxicity is usually suggested by symptoms that may include otalgia, tinnitus, and subjective hearing loss. In mild cases the main symptom is tinnitus, which often resolves within 1 week of cessation of therapy. The hearing loss may not appear until several days after treatment. Although it may be asymmetric, typically it is a bilateral sensorineural loss first at the higher frequencies and progressing into the lower frequencies with continued cisplatin therapy. Ototoxicity appears to be related more closely to the dose of the individual cycle rather than the total dose. However, there is individual susceptibility to cisplatin ototoxicity such that dramatic hearing deterioration can occur after a single high dose. Young children and the elderly are at higher risk, as are patients who concurrently receive either loop diuretics or aminoglycosides.

Cisplatin ototoxicity manifests histopathologically as outer hair cell loss most severe in the basal turn. The stria vascularis appears to undergo damage with continued toxicity, resulting in a decrease of the endocochlear potential. Pathological changes associated with cisplatin-induced vestibular toxicity remain to be established. The cellular mechanism of cisplatin-induced ototoxicity is probably multifactorial.

Cisplatin ototoxicity appears to be mediated, at least in part, by the production of free radicals, which results in the depletion of glutalhione and the antioxidant enzymes, such as superoxide dismutase, catalase, glutalhione peroxidase, and glutalhione reduction. The weakening of these antioxidant defense mechanisms in the cochlea results in the production of by-products, such as malondialdehyde and 4-hydroxynonenal. The latter has been shown to cause cell death by apoptosis in organ of Corti explants.

Several lines of evidence suggest that reactive oxygen species (ROS), or free radicals, are involved in the ototoxic side effects of cisplatin. A recent study by Clerici and colleagues demonstrated that ROS are generated in the cochlea of guinea pigs treated with ototoxic agents including cisplatin. ROS-mediated damage occurs in cisplatin-induced ototoxicity as a consequence of antioxidant depletion and increased lipid peroxidation in the cochlea of rats. The increases in superoxide anion and hydrogen peroxide have been reported to induce calcium-ion influx and pathological changes in cochlear cells. Administration of superoxide dismutase, glutalhione peroxidase, glutathione and its esters, and other antioxidants has also been shown to ameliorate cisplatin-induced nephro- and neurotoxicity in various species of animals. These studies indirectly provide the evidence that oxidative damage may be caused by cisplatin to the cochlea of animals.

The reduction of antioxidant enzyme system in the cochlea by cisplatin can increase the production of superoxide anion, hydrogen peroxide, and lipid peroxides. The imbalance between antioxidant protection and pro-oxidant reactions by cisplatin can result in the formation of toxic aldehyde by-products of lipid peroxidation. Huang et al. have shown in organ of Corti explants that 4-hydroxynonenal is produced following exposure to cisplatin. This aldehyde is extremely toxic and has been shown to cause cell death by apoptosis. This can lead to calcium in influx and pathological damage in outer hair cells.

Experimentally, these ototoxic effects of cisplatin can be reduced or eliminated by the administration of agents such as D-methionine, diethyldithiocarbamate (DDTC), methylthiobenzoic acid (MTBA), ebselen, and lipoic acid. These agents prevent the cisplatin-induced loss of outer hair cells in the cochlea, prevent auditory brainstem response (ABR), threshold shifts, and maintain normal levels of glutalhione and the antioxidant enzymes in the cochlea. They also prevent the increase in malondialdehyde (MDA) that occurs following cisplatin administration in experimental animals.

Another strategy to protect the cochlea is the use of inhibitors of cell death. Liu et al. have found excellent protection of the cochlea in vitro using inhibitors of the enzyme caspase, which is involved in the cascade that leads to cell death.

These findings in experimental animals suggest that depletion of antioxidant defense mechanisms in the cochlea leads to lipid peroxidation, which can trigger the enzymes leading to apoptosis, cell death, and hearing loss. It may be possible to use a combination of protective agents to prevent hearing loss and cell death of hair cells in the future if clinical trials show a beneficial effect without compromising the antitumor efficacy of cisplatin.

Carboplatin

Carboplatin is a highly effective antineoplastic agent that is active against small cell lung cancer as well as squamous cell carcinoma of the head and neck. It has also been effective in the treatment of malignant brain tumors when administered in combination with osmotic blood–brain barrier disruption (BBBD). Unfortunately, an unexpectedly high proportion of patients suffered high-frequency hearing loss following the treatment of brain tumors with this product.

Carboplatin is a unique ototoxic compound. Chinchillas treated with carboplatin have been reported to have selective destruction of the inner hair cells; the outer hair cells remain intact until high doses are reached. However, in the guinea pig, outer hair cells are selectively destroyed following carboplatin administration, whereas the inner hair cells are apparently less vulnerable.

A recent study by Doolittle et al. (2001) reported a marked difference in hearing loss following carboplatin treatment of patients with malignant brain tumors using a blood–brain barrier opening when sodium thiosulfate was administered as a protective agent 2 hours after carboplatin. Patients receiving carboplatin in conjunction with BBBD were found to have a 79% incidence of hearing loss. The average loss was 20 dB at 8 kHz after one treatment with carboplatin, and was more than 40 dB at that frequency after seven or eight treatments. Those patients who were protected with sodium thiosulfate lost only 3.7 ± 2 dB after one treatment. The sodium thiosulfate was given after the blood–brain barrier was allowed to close, thus preventing a direct reaction with carboplatin at the tumor site, which might otherwise inactivate the antitumor action of carboplatin.

The authors hypothesized that carboplatin may be inactivated by sodium thiosulfate in a manner similar to the inactivation of cisplatin. The thiol may bind to the platinum to form a complex that is rapidly excreted. The cochlea may perhaps concentrate sodium thiosulfate in a manner similar to that by which the kidney concentrates this thiol. This could provide protectant at the site of damage to prevent ototoxicity. Another possibility is that the inactivation occurs in the plasma.

A recent study suggests that carboplatin administration may result in free radical formation in the cochlea. Chinchillas that were treated with an intracochlear infusion of buthionine sulfoximine prior to carboplatin injection were found to have a significantly greater loss of both inner and outer hair cells than did chinchillas receiving carboplatin alone. Buthionine sulfoximine is an inhibitor of glutalhione synthesis. The finding that buthionine sulfoximine enhances carboplatin ototoxicity in animals provides support for the theory that free radicals and glutalhione play an important role in platinum ototoxicity. Thus depletion of glutalhione renders the cochlea more susceptible to free radical injury by carboplatin.

Salicylates

Sodium salicylate was used initially in 1875 for its antipyretic properties in the treatment of rheumatic fever. Aspirin ototoxicity occurs in 11 per 1000 patients, with much higher incidence reported for the long-acting aspirins. Elderly are at significantly higher risk of salicylate toxicity even at lower salicylate doses.

Salicylate is distributed throughout most body tissues, undergoes biotransformation by the liver, and is excreted by the kidney. Salicylates rapidly enter the cochlea after systemic administration, and the relationship between serum and perilymph salicylate level is linear. Manifestations of salicylate toxicity include nausea, vomiting, tinnitus, hearing loss, headache, mental dullness, confusion, tachycardia, and tachypnea. Hearing loss typically is flat, bilateral, and sensorineural. Loss can occur at salicylate concentrations below 30 mg/dL, is proportional to serum salicylate levels, and may result in threshold shifts of 30 to 40 dB. Complete reversibility occurs within 72 hours after cessation of the drug.

The auditory characteristics of salicylate-induced tinnitus have been characterized as tonal and high frequency. Pitch matching has identified involvement of tinnitus frequency 7 to 9 kHz. Tinnitus is the initial sign of salicylate ototoxicity and can be seen with salicylate levels as low as 19.6 mg/dL. Salicylate appears to alter spontaneous single-neuron activity, which may be the physiological manifestation of tinnitus.

The mechanism of salicylate ototoxicity appears to be multifactorial, likely secondary to reversible biochemical or metabolic changes in the cochlea rather than morphological abnormalities. Although there are no permanent histopathologic changes, the audiometric impairment is related to decreased cochlear blood flow or alterations in outer hair cell function. This is likely mediated by abnormal levels of arachidonic acid metabolites as well as, increased levels of catecholamines.

Quinine

Although the role of quinine as an antimalarial agent has been supplanted largely by less toxic semisynthetic derivatives, the increasing popularity of the drug for treatment of nocturnal leg cramps, as well as its use in tonic beverages, makes quinine ototoxicity a relevant clinical problem. Manifestations of quinine toxicity include hearing loss, vertigo, tinnitus, headache, nausea, and vision loss.

The sensorineural hearing loss typically is reversible, bilateral, and symmetric. It affects the high frequencies first with a characteristic 4 kHz notch and is often associated with poor speech discrimination.

Quinine extensively metabolizes in the liver, with only 10% being eliminated by the kidney unaltered.

Erythromycin

The macrolide antibiotic erythromycin was discovered in 1952, yet it has been only recently that its ototoxic potential has become known. As opposed to aminoglycoside hearing loss, which begins in high frequencies and can go unnoticed for some time, erythromycin-induced hearing loss occurs in speech frequencies as well as at high frequencies. Although this loss can be permanent, it often is reversible.

Risk factors for erythromycin ototoxicity include old age and hepatic or renal dysfunction. Caution should be exercised when administering erythromycin to patients who are receiving concomitant ototoxic medications. The mechanism by which erythromycin causes ototoxicity is not known.

Vancomycin

Vancomycin was introduced in the late 1950s as a breakthrough in the treatment of resistant *Staphylococcus aureus* infections, but it was soon replaced with the less toxic methicillin. In the 1980s vancomycin again received widespread acceptance for its use against methicillin-resistant *S. aureus* infections.

Vancomycin is often mistakenly identified as an ototoxic aminoglycoside antibiotic likely because of its *-mycin* suffix. Actually, vancomycin is a glycopeptide. An increased incidence of nephrotoxicity has been observed in humans receiving vancomycin and aminoglycosides in combination, and augmentation of ototoxicity has been demonstrated in experimental animals receiving vancomycin and aminoglycoside combination therapy. Although it is clear that vancomycin can greatly augment the ototoxicity of aminoglycoside antibiotics in experimental animals, there is no convincing evidence that it alone produces permanent ototoxicity.

AUDIOLOGICAL MONITORING FOR OTOTOXICITY

Audiological Test Methods Commonly Used in Ototoxicity Monitoring

Audiological monitoring for ototoxicity has been recommended for decades. Ototoxicity monitoring is currently designed to detect primarily cochleotoxic changes. Although tinnitus, vestibulotoxicity, and even central auditory changes can occur secondary to various ototoxins, they generally are not monitored, and no guidelines exist for them.

Early ototoxicity monitoring protocols depended primarily on pure-tone air conduction thresholds in the conventional audiometric testing range (250 Hz–8 kHz). Although this testing is still a part of most ototoxicity monitoring protocols, additional tests can greatly enhance patient care.

Bone conduction testing in the conventional frequency range, a standard part of most audiological assessments, has sometimes been ignored in ototoxicity monitoring other than at the baseline evaluation. Therefore, any audiometric threshold shift that occurred during or after drug administration was assumed to be sensorineural and indicative of ototoxic change. However, whenever a threshold shift is noted in the course of ototoxicity monitoring, the possibility of a conductive component must also be considered.

In some instances, these patient populations may be particularly susceptible to conductive hearing loss. For example, aminoglycoside patients may have infections (e.g., pneumonia) that increase the risk of otitis media. Chemotherapy patients may be immunosuppressed or exposed to head and neck radiation, both of which may increase the risk of middle ear effusion or infection.

Word recognition testing (i.e., the percent correct of words the patient can repeat at a suprathreshold listening level) is also important at baseline, and again any time a significant change in hearing threshold is noted. Word recognition is helpful in determining how clear speech is for the patient, at least in a quiet environment. Thus it is an essential component in designing patient management strategies when a hearing loss occurs. Although full 50-word lists with recorded stimuli generally are recommended in audiological assessments, they are particularly important in ototoxicity monitoring. With full 50-word lists, the Thornton and Raffin (1977) criteria for determining significant change can be applied.

High-frequency audiometry, developed specifically for ototoxicity monitoring, comprises air conduction thresholds from 10,000 to 20,000 Hz. High-frequency audiometry should be used routinely because most ototoxic hearing losses will first occur in that frequency region, which is higher than the frequency region required to understand speech. Therefore, this test can provide an "early warning sign" of ototoxic change before the patient incurs any change in the speech frequency range. Consequently, the physician and patient are given an opportunity to change the drug treatment protocol before permanent communicative impairment occurs. Without high-frequency audiometry, ototoxic change usually is first detected in the conventional or speech frequency range, greatly increasing the patient's risk of permanent disabling hearing loss. Unfortunately, some patients, particularly elderly and noise-exposed individuals, may not have sufficient hearing above 8000 Hz to allow high-frequency monitoring.

Otoacoustic emissions (OAEs) also have been demonstrated to show ototoxic changes prior to changes in the conventional test range (but currently there are no studies that have compared whether high-frequency audiometry or OAEs are better in first detecting ototoxic change). OAEs are acoustic signals, generated by the cochlea, that can be recorded by a small microphone in the ear canal. The most likely site for OAE generation is in the outer hair cells, one of the most vulnerable areas for many ototoxins. Consequently, it is logical that OAEs could provide an early indication of ototoxic change.

Spontaneous OAES occur even in the absence of acoustic stimuli and, when present, are considered a sign of good cochlear function. However, because they are present in under half of normal subjects, they have not been investigated for ototoxicity monitoring.

Both transient otoacoustic emissions (TOAEs), which occur in response to very brief acoustic stimuli (e.g., clicks), and distortion product otoacoustic emissions (DPOAEs), which occur in response to two simultaneous tones, have been reported to provide early indications of ototoxic change, before threshold shifts in the conventional frequency range. However, no study has yet compared TOAEs and DPOAEs to determine the relative values of these two measures. Therefore, TOAEs, DPOAEs, and high-frequency audiometry can be used to detect early ototoxic changes. Research is ongoing to determine the relative advantages and disadvantages of these three measures. However, OAEs can be recorded in very ill or even comatose patients and can quickly provide individual ear information in pediatric patients and thus are distinctly useful for those applications. OAEs are not always present in patients with preexisting hearing loss. Therefore, a thorough baseline evaluation can be used to determine which measures are appropriate.

ABR testing using high-frequency stimuli for ototoxicity monitoring is under investigation but has not yet been fully developed for clinical use.

More detailed information on audiological tests is available in Chapter 31.

Considerations in Selecting a Specific Monitoring Protocol for a Given Patient

Several factors must be considered in selecting an ototoxicity monitoring protocol, including the type and degree of ototoxic risk, the drug administration schedule, the patient's age and ability to cooperate with testing, the test environment, and whether or not the drug protocol can be modified if ototoxic change occurs.

The most commonly monitored drugs are platinum-based antineoplastics (e.g., cisplatin, carboplatin) and aminoglycoside antibiotics (e.g., amikacin, tobramycin, gentamicin).

Complete baseline evaluations prior to drug administration are essential. Preexisting hearing loss is common, and without a baseline evaluation, there is no reference for possible ototoxic change. The baseline evaluation should include pure-tone air and bone conduction audiometry in the conventional frequency range, speech reception threshold testing, word recognition testing, immitance audiometry if any conductive component is present, and high-frequency audiometry and/or OAEs. In many cases, it is advisable to use all measures, including OAEs and high-frequency audiometry, at baseline. This will ensure that if the monitoring protocol later needs to be modified (e.g., patient becomes unable to cooperate behaviorally), appropriate baseline information is still available. Periodic monitoring generally will include only the air conduction thresholds in the conventional range, and high-frequency audiometry and/or OAEs. However, if a change is noted, then the rest of the test battery included in the baseline assessment is added to determine if a conductive component exists and the impact of the change on word recognition.

The platinum compounds are usually delivered once every 3 to 4 weeks, and audiological monitoring is recommended just prior to the start of each treatment. Although platinum-induced hearing loss is usually permanent, this timing allows for any transient shift from the previous treatment to resolve. This schedule also allows these patients to be tested when they are most able to concentrate on the listening tasks and before they are connected to noise-generating equipment.

Aminoglycoside antibiotics generally are administered daily. Monitoring is recommended 1 to 3 times per week. Both platinum compounds and aminoglycoside antibiotics can cause delayed hearing loss. Therefore a post-treatment audiological assessment at 3 to 6 months is recommended.

Because ototoxins can render the patient more susceptible to noise-induced hearing loss, patients should be advised to avoid noise exposure during and for a few months after treatment. They also may need to be reminded to apprise their physician and audiologist of all medications they take, even over-the-counter medications, during this period. For example, self-administered high-dose aspirin could cause misleading findings.

An individual patient's ability to cooperate with testing may play a role in which tests are selected. Most adult patients will be able to cooperate with a full range of tests, including pure-tone air and bone conduction testing for the conventional frequency range, high-frequency audiometry, and OAEs. However, some patients may be too ill to tolerate lengthy testing or may be unresponsive. In these cases, OAEs and ABRs, which do not require a patient response, may be useful. Additionally, some investigators recommend only a limited high-frequency testing range for monitoring, followed by full testing if a change occurs.

Bedside testing can be performed if the patient is too ill to be transported to the clinic. However, bedside testing is not the first choice because the ambient electrical and acoustic noise can interfere with accurate threshold assessment. Also, a patient who is too ill to be transported to the clinic is frequently too ill to provide reliable threshold responses. In these cases, OAE monitoring may be appropriate, but only if OAE baseline information is available.

Preschool children or infants, particularly when they are ill, may not provide reliable behavioral thresholds under earphones. Behavioral thresholds may be limited to the sound field environment, which is useful but does not provide individual ear information. Therefore, OAEs are generally an essential component of ototoxicity monitoring in children. OAEs can provide excellent individual ear information of cochlear function and an early warning of ototoxic change. Additionally, they can be measured in less than 5 minutes per ear, which is a distinct advantage for pediatric patients.

In some cases, the patient's drug treatment protocol cannot be modified even if ototoxicity occurs. In those cases, audiological monitoring is still recommended to help patients and their family deal with any communication deficit that occurs. Maintaining communication is particularly important for patients and their families during serious, life-threatening illness. However, if the drug treatment protocol cannot be altered, regardless of audiological findings, high-frequency audiometry and/or OAEs to provide early detection of ototoxic changes may not be necessary. Instead, monitoring within the conventional frequency range may be adequate to determine when the patient requires assistance with communication.

In most cases, the earliest information possible on ototoxic changes is needed to allow the patient and physician to consider alternative treatment options before irreversible hearing loss in the speech frequency range occurs. Therefore, high-frequency audiometry and/or OAEs are almost always included in ototoxicity monitoring protocols in addition to monitoring in the conventional frequency range.

Determination of Significant Ototoxic Change

For ototoxicity monitoring, all test results should be compared with the baseline evaluation rather than with results from the most recent test. Often ototoxic change occurs very gradually and can be missed if not compared with initial baseline.

Significant ototoxic change in the conventional frequency range generally is determined according to the guidelines of the American Speech-Language-Hearing Association (1994). These criteria are ≥20 dB decrease at any one test frequency, ≥10 dB at any two adjacent test frequencies, or loss of response at three consecutive test frequencies where responses were previously obtained. The latter criterion refers specifically to the highest frequencies tested, where earlier responses are obtained close to the limits of audiometric output and later responses cannot be obtained at the limits of the audiometer. Changes must be confirmed by repeat testing.

The same criteria could be applied to high-frequency audiometry, although no well-established guidelines exist. Other recommendations for significant change criteria for high-frequency audiometry include threshold change >10 dB at 8000 to 14,000 Hz or >15 dB at 16,000 to 18,000 Hz, which is similar to suggestions by Frank et al (1990). Other recommendations include ≥20 dB shift at one frequency, ≥15 dB at two frequencies, or ≥10 dB at four or more frequencies across the conventional and high-frequency ranges. Additionally, different criteria may be needed for abbreviated high-frequency audiometry protocols.

No well-established criteria exist for ototoxic changes in OAEs, although several studies have documented that OAEs can diminish or disappear prior to ototoxic change in the conventional frequency range. Clinically, for OAEs, the absence of a replicable response in a frequency band that had a response at baseline is often considered a significant change. Absolute amplitude measures can be variable from test to retest and may not be the best criteria to employ. However, further research is continuing in this area, and more specific criteria may be developed.

Audiological Management of Ototoxicity

The audiologist should counsel the patient on the initial visit regarding the importance of and reasons for monitoring, along with the monitoring schedule. The patient also should be advised to avoid noise exposure and to report all other medications during and usually for a few months after the monitoring period. The patient should be reassured that every effort is being made to detect any possible changes as early as possible to avoid any permanent communicative handicap. But the patient needs to be further assured that even if permanent hearing loss occurs, assistance will be available.

Test results, particularly those showing a significant change, must be sent to the physician on the same day of testing. Often, the patient is scheduled for drug administration immediately after the monitoring session, and in these cases a note should be sent with the patient, or the physician should be telephoned. Rapid notification is essential to allow the physician to change the drug protocol, if possible, before further hearing change occurs.

Even with careful monitoring, sometimes permanent hearing loss does occur that impairs communication either because the drug protocol could not be changed or because there was a sudden and sometimes severe onset of ototoxicity. In these cases, management is similar to that for other patients with sensorineural hearing loss. However if the patient is fit with hearing aids, the gain and maximum power output are kept as low as possible to avoid any noise exposure that could be synergistic with the ototoxin. The output can be readjusted a few months after the ototoxin is discontinued. Also, some terminally ill patients may choose amplification even for a relatively short time because they wish to maintain communication with their families and handle their affairs. Counseling the patient's family regarding communication strategies can also be helpful.

The primary concern is always to maintain the patient's communication capabilities during what is generally a serious illness.

Suggested Readings

American Speech-Language-Hearing Association (ASHA) (1994). Guidelines for the Audiologic Management of Individuals Receiving Cochleotoxic Drug Therapy. ASHA 36 (Suppl. 12): 11–19

Campbell KCM, Durrant J. Audiologic monitoring for ototoxicity. Otolaryngol Clin NA 1993;25(5):903–914

Chiodo AA, Alberti PW. Experimental, clinical and preventive aspects of ototoxicity. Eur Arch Otolaryngology 1993;16:141

Clerici WJ, Hensley K, DiMartino DL, et al. Direct detection of ototoxicant-induced reactive oxygen species generation in cochlear explants. Hear Res. 1996;98:116–124

Doolittle ND, Muldoon LL, Brummett RE, et al. Delayed sodium thiosulfate as an otoprotectant against carboplatin-induced hearing loss in patients with malignant brain tumors. Clin Cancer Res 2001;7:493–500

Dreschler WA, van der Hulst RJAM, Tnage RA, et al. Role of high-frequency audiometry in the early detection of ototoxicity, I: Clinical aspects. Audiology 1989;28:211–220

Frank T. High Frequency Hearing Thresholds in Young Adults Using a Commercially Available Audiometer. Ear Hear 1990;11: 450–454

Huang T, Cheng AG, Stupak H, et al. Oxidative stress-induced apoptosis of cochlear sensory cells: otoprotective strategies. Int J Dev Neurosci. 2000;18:259–270

Koegel CJ. Ototoxicity: a contemporary review of aminoglycosides, loop diuretics, acetylsalicylic acid, quinine, erythromycin, and cisplatin. Am J Otol 1985;6:190–199

Littman TA, Magruder A, Strother DR. Monitoring and predicting ototoxic damage using distortion-product otoacoustic emissions: pediatric case study. J Am Acad Audiol 1998;9:257–262.

Liu W, Staecker H, Stupak H, et al. Caspase inhibitors prevent cisplatin-induced apoptosis of auditory sensory cells. Neuroreport 1998;9:2609–2614

Mills RR. Deafness due to plain and long-acting aspirin tablets. J Clin Parmacol 1978;18:468–471

Smith CV. Free radical mechanisms of tissue injury. In: Mosler MT, Smith CV, eds. Free Radicals of Mechanisms of Tissue Injury. Boca Raton, FL: CRC Press; 1992:1–22

Thornton A, Raffin MJM. Speech Discrimination Scores Modeled as a Binomial Variable. J. Speech Hear Res 1977;21:507–518

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. Which of the following drugs is most likely to cause permanent hearing loss?
   A. erythromycin
   B. quinine
   C. cisplatin
   D. penicillin G
2. High-dose salicylate treatment can cause
   A. a transient hearing loss that is reversible after the drug is stopped
   B. permanent hearing loss that does not recover after the drug is stopped
   C. permanent damage to the vestibular system that does not recover after the drug is stopped
   D. transient dizziness that stops after the drug is stopped
3. The use of a loop diuretic in combination with an aminoglycoside antibiotic
   A. can potentiate the antimicrobial effects of the antibiotic without causing any additional ototoxicity
   B. can cause a temporary hearing threshold shift that recovers after the drugs are stopped
   C. exacerbates the ototoxic effect of the aminoglycoside antibiotic, resulting in permanent hearing loss
   D. has no effect on hearing

Chapter 10A

ONCOLOGY OF HEAD AND NECK TUMORS

ELIZABETH FRANZMANN, SCOTT LILLY, DAVID HUANG, AND GIOVANA THOMAS

The anatomy, proximity of vital structures, and variety of tissues within the head and neck contribute to the complexities in the management of cancers of this area. The morbidity associated with head and neck neoplasms and their treatments is disproportionately high due to their effect on speech, taste, smell, deglutination, and cosmetics. Head and neck malignancies comprise a diverse group including lymphoma, sarcomas of bone and soft tissue, carcinomas of the salivary and thyroid glands, and multiple skin cancers such as basal cell, squamous cell, and melanoma. Head and neck squamous cell carcinoma (HNSCC) accounts for 95% of all malignancies in the upper aerodigestive tract (UADT) and will be the focus of this chapter.

Despite marked improvement in oncology therapies over the past several decades, the survival for HNSCC has improved by less than 10% in the United States. The majority of head and neck cancers are locoregionally advanced at the time of initial diagnosis, and only 50% are cured. Head and neck oncologists are investigating better prevention, early detection, and treatment strategies.

EPIDEMIOLOGY

According to the American Cancer Society's *Facts and Figures* (2003), in the United States cancers of the oral cavity, pharynx, and larynx accounted for nearly 3% of incident cancers and 2% of cancer deaths in the U.S. Men are affected more than twice as often as women. Over half of these cancers involve the oral cavity. The rest are divided equally between the larynx and pharynx.

Worldwide, head and neck cancer accounts for over a half million new cases of cancer annually.

RISK FACTORS

The most important known predisposing factors for the development of HNSCC are tobacco and alcohol use. Repeated prolonged exposure of the mucosa of the aerodigestive tract to such carcinogens results in field cancerization leading to multiple independent malignant squamous-type epithelial neoplasms that may develop either concomitantly or in sequence. Depending on the type of tobacco product used, HNSCC site studied, and amount of exposure, risk to smokers can be 40 times that of nonsmokers. Alcohol use appears to be an independent risk factor and acts synergistically with tobacco use in causing HNSCC. In India and Southeast Asia, chronic use of the beetle quid (paan) is strongly associated with increased risk of HNSCC.

Susceptibility and gene–environment interactions may play a role in HNSCC. Only a fraction of individuals exposed to tobacco and alcohol use will develop the disease. Evidence suggests that a family history of cancer may be a risk factor. Polymorphisms in genes that affect carcinogen metabolism have been implicated in HNSCC risk. Deoxyribonucleic acide (DNA) repair capacity is important to normal cell function, and variants of several genes involved in DNA repair have been associated with increased risk of HNSCC.

Because approximately 15% of HNSCC cases occur in nondrinkers and nonsmokers, other risk factors have been investigated. None show as great a risk as tobacco and alcohol use. Purported risk factors that have been discussed include environmental tobacco smoke, human papillomavirus, poor diet, certain occupations and occupational exposures, low education and socioeconomic status, immunodeficiency, mouthwash use, poor oral hygiene, gastroesophageal reflux disease, and a benign oral condition known as oral lichen planus.

DYSPLASIA

There is good evidence that UADT mucosa progresses through a premalignant phase prior to development of frank malignancy. Dysplasia is a premalignant lesion diagnosed histologically. Dysplastic lesions transform to carcinoma 6 to 36% of the time, depending on severity. Dysplasia most often occurs after the fifth decade of life, though the average age may be declining. It is identified when white (leukoplakia), red (erythroplakia), or mixed (erythroleukoplakia) lesions are seen on physical exam and biopsied.

MULTIPLE PRIMARY TUMORS, REGIONAL METASTASES, AND RECURRENCE

Because the entire mucosal surface of the UADT is exposed to carcinogens, multiple sites are at risk for developing HNSCC. The chance of developing a second primary tumor in patients with HNSCC has been estimated at 2 to 3% per year.

HNSCC tends to spread regionally to involve cervical lymph nodes. The most significant risk factors for regional spread are tumor site and size, grade of histological differentiation, tumor thickness (tongue and floor of mouth carcinoma), vascular embolization, and perineural infiltration.

HNSCC has a propensity to recur. A recurrent tumor at the primary site occurs in 20 to 30% of patients and is the most common cause of treatment failure. Regional recurrence in the neck occurs in 10 to 15% of patients and is the most common cause of disease-related death in these patients. Prognosis is bleak with late-stage recurrence; early diagnosis is imperative.

EVALUATION OF THE HNSCC PATIENT

A primary care physician or dentist usually diagnoses HNSCC. The majority of patients present in late stage and are referred to a head and neck specialist for treatment. After performing a complete history and physical examination, the surgeon verifies the pathologic diagnosis and orders additional imaging to complete the staging process. This usually includes a computed tomographic (CT) scan of the neck, a chest x-ray, and liver function tests. If the patient has extensive (large or diffuse) cervical adenopathy, a CT scan of the chest is preferred over a chest x-ray. Other metastatic workup [i.e., bone scan, positron emission tomography (PET) scan, brain CT] can be performed as indicated by the patient's signs and symptoms. A thorough examination for second primary cancer should be performed. This is usually done in the operating room with direct laryngoscopy, bronchoscopy, and esophagoscopy. The patient's airway and breathing should be evaluated and a tracheostomy performed if the airway is tenuous.

The patient's nutritional status is evaluated and supplemented as necessary to optimize healing potential. If the patient is unable to swallow, a feeding tube is placed. The patient should undergo a dental evaluation. Concurrent medical problems such as lung and heart disease are frequent, given the risk factors for this disease. Because treatment modalities are rigorous, other medical issues should be evaluated expeditiously and treated by appropriate specialists so that the patient's functional status is optimal prior to treatment.

PROGNOSIS

Prognosis varies with site and stage of lesion. In general, patients with laryngeal cancer have the best prognosis. Patients with hypopharyngeal cancer have the worst prognosis. Head and neck cancer is classified as stages I to IV based on the AJCC TNM (tumor, node, metastasis) staging system (**Tables 10A–1** to **10A–6**). Classification is anatomical site specific: oral cavity, oropharynx, nasopharynx, hypopharynx, supraglottis, and glottis. Tumor stage depends on site, but generally T1 to T3 indicates increasing tumor size, and T4 indicates extension to adjacent sites. Node staging is based on size, number, and side of neck involved and is uniform for all sites except the nasopharynx. Metastasis (M) is either MO (no metastasis), MI (metastatic), or MX (unknown). The risk of distant metastases (stage IV) increases with increasing neck disease. Patient cure is achieved in over 80% of stage I patients and over 60% of stage II patients. For patients with more advanced disease (stages III and IV), cure is attained in fewer than 30% of cases, although some stage III patients have a better prognosis.

TABLE 10A–1 THE TNM SYSTEM OF A PRIMARY TUMOR*

Tumor stage	*Definitions*
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ
Metastasis stage	
M0	No metastasis
M1	Metastasis
Stage grouping for all head and neck sites (except nasopharynx)	
Stage I	T1N0M0
Stage II	T2N0M0
Stage III	T3N0M0
	T1, T2, or T3N1M0
Stage IVA	T4N0 or N1M0
	Any T N2M0
Stage IVB	Any T N3M0
Stage IVC	Any T, any NM1

*Other tumor classes vary by anatomical site.
TNM, tumor, node, metastasis.

TABLE 10A–2 ORAL CAVITY AND OROPHARYNX TUMOR STAGES

T1	Tumor <2 cm in greatest dimension
T2	Tumor >2 cm but not >4 cm in greatest dimension
T3	Tumor >4 cm in greatest dimension
T4	Tumor invades adjacent structures [e.g., cortical bone, soft tissue of neck, deep (extrinsic) muscle of tongue]

TREATMENT

Treatment regimens vary with HNSCC tumor site, stage, and institutional preference. Generally, early (stage I or II) lesions can be treated with surgery or radiation alone. Late-stage lesions are treated with combinations of surgery, radiation, and chemotherapy.

SURGICAL CONSIDERATIONS

The patient's functional status and preferences, prior treatment, and the tumor size and site all affect surgical decision making. These factors should be weighed carefully to determine the best treatment approach.

Oral Cavity

The oral cavity is bounded anteriorly by the junction of skin and vermilion border of lip, posteriorly by the hard and soft palate junction, circumvallate papillae, and anterior tonsillar pillars, and laterally by the buccal mucosa. Subsites within the oral cavity include the alveolus, retromolar trigone, floor of the mouth, oral tongue, lip, and hard palate. HNSCC of the oral cavity is particularly amenable to surgery. Smaller tumors are easily excised

TABLE 10A–3 SUPRAGLOTTIS TUMOR STAGES

T1	Tumor limited to one subsite of supraglottis with normal vocal cord mobility
T2	Tumor invades more than one subsite of supraglottis or glottis, or region outside the supraglottis (e.g., medial wall of piriform or mucosa of tongue base) without fixation of the larynx
T3	Tumor limited to larynx with vocal cord fixation or invades postcricoid area or pre-epiglottic tissues
T4	Tumor invades thyroid cartilage, or extends to soft tissues of the neck, thyroid, or esophagus

TABLE 10A–4 GLOTTIS TUMOR STAGES

T1	Tumor limited to vocal cord(s) (may involve anterior or posterior commissures) with normal mobility
T1a	Tumor limited to one vocal cord
T1b	Tumor involves both vocal cords
T2	Tumor extends to supraglottis or subglottis, or with impaired vocal cord mobility, or any combination of these
T3	Tumor limited to the larynx with vocal cord fixation
T4	Tumor invades thyroid cartilage or extends to other tissues beyond the larynx (e.g., oropharynx, soft tissues of neck)

with little morbidity. Larger lesions tend to invade bone and are better cleared with surgery followed by radiation. If depth of invasion is significant, treatment of the neck is recommended. Depending on size and site, defects can be left to granulate, closed primarily, or covered with a skin graft, a local flap, a locoregional flap, or a free flap. For lesions that involve the periosteum of the mandible with minimal cortical invasion, a marginal mandibulectomy usually will suffice, given that there is adequate mandibular height. For more significant invasion, a segmental madibulectomy is recommended. Lateral mandibular defects can be reconstructed with a pectoralis major mycutaneous flap, whereas anterior defects are better closed with a fibular or iliac crest free flap. (For tumor classification, see **Table 10A–2**.)

Oropharynx

The oropharynx lies posterior to the oral cavity and extends from the level of the hard palate to the hyoid bone. It is bounded anteriorly by the junction of the hard and soft palates, anterior tonsillar pillars, and circumvallate papillae. The oropharynx includes the base of the tongue, tonsils, soft palate, and posterior pharyngeal wall.

TABLE 10A–5 HYPOPHARYNX TUMOR STAGES

T1	Tumor limited to one subsite of hypopharynx and ≤2 cm in greatest dimension
T2	Tumor invades more than one subsite of hypopharynx or an adjacent site, or measures >2 cm but not >4 cm in greatest diameter without fixation of hemilarynx
T3	Tumor measures >4 cm in greatest dimension or with fixation of the hemilarynx
T4	Tumor invades adjacent structures (e.g., cartilage or soft tissues of neck)

TABLE 10A–6 NASOPHARYNX TUMOR STAGES

T1	Tumor confined to nasopharynx
T2	Tumor extends to soft tissues of oropharynx or nasal fossa
T2a	Without parapharyngeal extension
T2b	With parapharyngeal extension
T3	Tumor invades bony structures or paranasal sinuses
T4	Tumor with intracranial extension or involvement of cranial nerves, infratemporal fossa, hypopharynx, or orbit

Cancers of the oropharynx can be treated surgically. Common approaches to the oropharynx include suprahyoid, mandibulotomy, and lateral pharyngotomy. Very small lesions can be resected transorally. Large lesions involving the mandible require composite resection. A pectoralis major myocutaneous flap provides adequate coverage of soft tissue defects in this region, although a radial forearm free flap tends to be thinner, can be reinnervated, and may allow better swallowing function. For large defects or if the patient has had previous radiation, soft tissue coverage with a vascularized flap is recommended to avoid fistula and vessel exposure. (For tumor classification, see **Table 10A–2**.)

Larynx

The larynx is divided into supraglottic, glottic, and subglottic regions. The supraglottic larynx includes the epiglottis, aryepiglottic folds, false vocal cords, and roof of the ventricle. The glottic larynx includes the floor of the ventricles, true vocal cords, and anterior and posterior commissures. The subglottis begins 0.5 mm below the free edge of the vocal cords and extends to the inferior border of the cricoid cartilage.

Surgical treatment of laryngeal cancer is diverse. Surgical options include conventional conservation surgery, transoral endoscopic laser surgery, supracricoid partial laryngectomy, and total laryngectomy. The theory behind conventional conservation surgery is based on work by Kirchner that shows the spread of cancer is influenced by fibroelastic ligaments and membranes, which confine the tumor to anatomical compartments.

Conventional conservation surgery can be divided into vertical and horizontal approaches. Vertical approaches involve anterior and posterior vertical incisions that include a segment of supraglottis, glottis, and subglottis. Operations under this classification include the hemilaryngectomy and frontolateral laryngectomy. Horizontal approaches involve superior and inferior horizontal incisions that remove lesions above the true vocal cords. Operations under this classification include supraglottic laryngectomy and extended supraglottic laryngectomy. The supracricoid laryngectomy expands indications for conservation surgery. It involves removal of a cylindrical segment of larynx above the cricoid, maintaining the arytenoids on at least one side. Limitations include arytenoid fixation, posterior commissure involvement, cricoid cartilage involvement, and extralaryngeal spread.

For laryngeal tumors that invade through cartilage, strap muscles, or significantly into the subglottis, conservation surgery is not appropriate, and organ-sparing chemoradiation protocols are not as effective. For these very advanced laryngeal tumors, total laryngectomy is the treatment of choice. (For tumor classification, see **Tables 10A–3** and **10A–4**.)

Hypopharynx

The hypopharynx partially surrounds the larynx. It begins at the level of the hyoid bone and ends at the esophageal introitus. The hypopharynx includes the piriform sinuses, pharyngeal walls, and postcricoid region.

HNSCC in the hypopharynx tends to behave aggressively. Squamous cell cancers in the hypopharynx that are not cured following organ-sparing chemotherapy and radiation are often difficult to resect or are unresectable. If the tumor is resectable at the time of presentation, many oncologists recommend primary surgery followed by radiation. Surgery involves laryngectomy and partial pharyngectomy and often requires closure with vascularized tissue such as a pectoralis major myocutaneous flap. Obviously, the patient's preferences in such a decision are critical. If risks are understood, organ-sparing protocols are reasonable options as well.

Nasopharynx

The nasopharynx is bounded superiorly by the basiocciput and basisphenoid, posteriorly by the C1 and C2 cervical bodies, anteriorly by the choanae, and inferiorly by the soft palate. The lateral walls are occupied primarily by the eustachian tube orifice. Immediately posterior to the eustachian tube orifice is the fossa of Rosenmüller, where most nasopharyngeal carcinomas originate. Due to the close proximity of cranial nerves, nasopharyngeal carcinoma often presents with cranial nerve dysfunction. The most common cranial nerves involved are VI and V; however, cranial nerves II, IV, IX, X, XI, and XII can be involved. Most nasopharyngeal carcinomas respond well to radiation and chemotherapy. (For tumor classification, see **Table 10A–6**.)

Neck

Lymph node groups are divided into levels. Level 1 includes the submandibular and submental nodes, levels 2 to 4 are located between the posterior border of the sternocleidomastoid (SCM) and the strap muscles. These are separated into upper jugular (level 2), midjugular (level 3), and lower jugular (level 4) nodes. Level 5 is located behind the SCM and extends to the trapezius. It is bordered inferiorly by the clavicle.

Neck dissections performed for HNSCC include the radical neck dissection, modified radical neck dissection, and selective neck dissections. The radical neck dissection includes removal of all five nodal groups, the SCM, the internal jugular vein (IJV), and the spinal accessory nerve (CN11). The modified neck dissection involves removal of all five nodal groups; however, one or more of the other three structures (SCM, IJV, or CN11) are preserved. Selective neck dissections remove only those nodal groups at highest risk for tumor involvement. The risk depends on primary tumor location and size.

Generally, a modified radical neck dissection is performed on all clinically positive neck disease if the primary site is also being managed surgically. Neck dissections are also performed on all persistent neck disease following definitive organ-sparing protocols. At some institutions, elective neck dissections are performed on patients with nodal disease greater than 3 cm prior to chemoradiotherapy. Others recommend neck dissection for all N2 disease. Still others recommend observation if the neck is clinically negative following treatment. Neck dissections generally are recommended in patients with primaries greater than 4 mm in depth in the oral cavity, for patients with supraglottic cancer, and for patients with any advanced HNSCC. For supraglottic tumors, bilateral neck dissections are performed because lymphatics cross in the midline. Increasingly, selective neck dissections are performed in the N0 neck. Levels 1 to 3 (or 4) are dissected for oral cavity cancers and levels 2 to 4 for laryngeal, oropharyngeal, and hypopharyngeal primary sites.

Radiotherapy

Radiotherapy may be used to treat head and neck cancers as the definitive (curative) therapy or as the postoperative (adjuvant) therapy. When radiotherapy is used for the definitive purposes, it may be given as a single modality or in a combined modality with chemotherapy. Recent advances include the use of intensity modulated radiotherapy (IMRT), which arranges the radiation beam so there is a sharper dose gradient between the target tumor volumes and the normal tissue/organs. IMRT provides dose escalation to the tumors and additional normal tissue sparing. The ability to deliver high doses to a selected volume enables the radiation oncologist to reirradiate recurrent diseases/second primaries in a previously irradiated field. In the following sections, brief discussions of the role of radiotherapy in the management of head and neck cancers will be presented.

Single Modality

Surgery is indicated for early-stage diseases of the head and neck region if there is no sacrifice of a critical normal organ function or if the patient has no contraindication to surgery. Prognosis for early-stage diseases (AJCC stage groupings I and II) depends on the anatomical site, the size of the tumors, the nodal status, the histology/grade of the diseases, and the patient's physical condition, such as weight loss and the ability to tolerate the treatment. In general, the radiation dose ranges from 66 to 72 Gy with a 1.8 to 2.0 Gy daily dose. Other factors that affect the success of radiotherapy are the total dose delivered, the daily dose, and the frequency of treatment interruptions (i.e., the number of elapsed days to complete a course of radiotherapy). Typical control rate for a T1 tonsillar lesion is approximately 80 to 85%; for T2 tonsillar lesions, 75 to 85%. Early-stage lesions in the larynx have higher control rates than other head and neck sites if treated with radiotherapy alone. The control rate for T1a glottic cancers is approximately 90%; for T1b tumors, approximately 80%. The laryngeal function is preserved in 74% of T2 glottic cancers.

Hyperfractionated radiotherapy, which delivers a dose twice daily, is used to prevent late-occurring side effects such as fibrosis and nerve damage. Generally, doses are 1.2 Gy twice daily. This technique was studied by the University of Florida with excellent results for treatment of T3/T4 laryngeal lesions (65% local control for T3 and 60% for T4).

Other altered fractionations, such as accelerated concomitant boost (accelerated hyperfractionation), which uses two different doses in the last 10 to 12 days of radiotherapy, and continuous, hyperfractionated, accelerated radiotherapy (CHART, three doses per day) are adopted for overcoming the repopulation effect after initiation of radiotherapy with or without chemotherapy. The commonly used doses for the accelerated hyperfractionation are 1.8 Gy and 1.5 Gy. The CHART regimen uses 1.5 Gy 3 times daily.

A randomized study, RTOG 9003 (Radiation Therapy Oncology Group, Fu et al, 2000), showed that the accelerated hyperfractionation regimens improve local tumor control significantly, by approximately 10 to 15%. There is a trend toward better disease-free survival as compared with standard daily radiotherapy. The CHART regimen demonstrated the importance of controlling tumor repopulation to prevent tumor recurrence. It was also observed that the acute radiation side effects were more severe among patients who received the accelerated hyerfractionation regimen than patients who were treated with standard fraction.

Chemoradiotherapy

For locally advanced diseases; that is, stages III and IV, the treatment results for radiotherapy alone are not as successful as surgery with or without postoperative radiotherapy. The Veterans Affairs Laryngeal Cancer Study Group trial in 1991 was a landmark investigation, that introduced chemotherapy as a standard treatment option for advanced laryngeal cancer. The study showed that induction chemotherapy using *cis*-platinum and 5-Fluorouracil (5-Fu) followed by radiation therapy, with surgical salvage of tumors that were unresponsive, resulted in preservation of the larynx in over 60% of patients while maintaining survival rates equal to total laryngectomy followed by radiation. This approach quickly became the standard of care for appropriate patients who wanted to preserve their larynx.

Recent randomized phase III trials of cisplatin-based induction chemotherapy followed by radiotherapy with or without concurrent chemotherapy for laryngeal and hypopharyngeal cancers showed that organ preservation can be achieved with similar overall survival as surgery followed with radiotherapy. Concurrent chemoradiotherapy is superior to induction chemotherapy followed by radiotherapy.

Chemotherapy that was given alternately with radiotherapy was also shown in a randomized phase III study to improve significantly overall survival for stage III and IV head and neck cancers. An RTOG study (H 00129) is ongoing to compare concurrent chemotherapy with daily radiotherapy (2 Gy daily) and accelerated hyperfraction radiotherapy (1.8 Gy + 1.5 Gy in the accelerated phase) for the diseases in the oropharynx and oral cavity. Most practices are using combined modality chemoradiotherapy with cisplatin and 5-fu for locally advanced stage head and neck cancers, even with limited data.

Adjuvant Radiotherapy

Radiotherapy is often indicated postoperatively (adjuvant therapy). It is commonly recommended for positive margins, large tumors, extracapsular extension of lymph node metastasis, more than three involved lymph nodes, multiple recurrences for high-grade lesions, perineural invasion, or situations in which the operating surgeon is concerned for residual diseases. Postoperative radiotherapy has been shown to improve survival. A treatment-related factor that affects local tumor control is the elapsed time between surgery and the beginning of postoperative radiotherapy. Although recently debated, most studies agree that the local recurrence rate is significantly higher if radiotherapy begins more than 6 weeks after surgery.

Reirradiation

When previously irradiated patients develop in-field recurrence or a second malignancy, surgical salvage may not always be feasible. In present practice, chemotherapy may be offered to patients for palliative treatment. Reports from the University of Texas M. D. Anderson Cancer Center as well as Massachussetts General Hospital showed that for selected patients treated again with chemotherapy and reirradiation, the reported local control rate may be as high as 50%, and the 5-year survival may be up to 20%. The treatment-related side effects are well tolerated. RTOG recently closed a phase II study (RTOG 99-11) to evaluate the efficacy of this approach. The treatment regimen is 1.5 Gy twice daily every other week to a total dose of 60 Gy. Cisplatin and Paclitaxel are given daily with radiotherapy. Radiation myelitis is expected to be less than 6%. A review of the toxicity of the enrolled cases showed no excessive number of carotid artery hemorrhages.

Management of Xerostomia and Mucositis

The acute radiation reactions that have the most impact on a patient's sense of well-being are xerostomia and oral mucositis. For xerostomia, artificial saliva may be used. Pilocarpine, a cholinomimetic drug used to increase salivation, has been given frequently to patients by radiation oncologists with mixed results. The benefit from the use of oral pilocarpine is still controversial. Amifostine, a prodrug analogue of cysteamine thought to protect normal cells by acting as a free radical scavenger, may be used alone or together with pilocarpine. The combined use of these two drugs is being investigated as well. For selected sites of head and neck cancers, transposition of submandibular gland to the submental region has shown effectiveness in single-institution experience. RTOG is conducting a study (RTOG 0244) to investigate the benefit of this procedure. Carefully planned IMRT for treating head and neck cancers has been shown to decrease the severity of xerostomia significantly without sacrificing tumor control.

The management of oral mucositis includes pain control and prevention of infection. Local anesthetics such as viscous xylocaine may be used. For severe pain, fentanyl patches, Roxicet elixir, and Roxanol may be added. Morphine mouthwash has been shown to be effective. Nystatin oral suspension for topical use with fluconazole for more serious infection is commonly practiced. A mixed solution that may contain a combination of Benadryl, Mylanta, Decadron, nystatin, antibiotic, morphine, or viscous xylocaine is also widely used for head and neck patients. Rinsing the mouth with saline or sodium bicarbonate solution frequently is a preventive measure. It is also important for patients to have frequent evaluations by a dentist with experience in treating patients with HNSCC. Special emphasis on hydration and nutrition is important to help patients tolerate chemoradiotherapy for head and neck cancers.

Chemotherapy

Chemotherapeutic agents are used in combination with surgery and radiation to treat advanced head and neck cancer. Such regimens include induction chemotherapy, concomitant chemoradiotherapy, adjuvant chemotherapy, and palliative chemotherapy.

Induction and Concomitant Chemotherapy

The Veterans Affairs study showed that induction chemotherapy and radiotherapy are equivalent to laryngectomy followed by radiotherapy. However, it did not include a radiation alone arm; hence the contribution of induction chemotherapy was unclear. More recently, the RTOG 91-11 compared induction chemotherapy followed by radiation, concurrent chemoradiation, and radiation alone. In this trial, 88% of patients in the concomitant chemoradiation arm preserved their larynx at 2 years compared with 75% of those treated with induction chemotherapy and 70% of those treated in the radiation-alone arm. Overall survival did not vary significantly between groups. Concomitant chemoradiation therapy is now the preferred approach to laryngeal preservation, although radiation can be advised for patients who cannot withstand the toxicity of combined chemoradiation.

Concomitant chemoradiotherapy has been postulated to improve locoregional control and eradicate systemic micrometastasis by the following mechanisms:

1. Spatial cooperation: radiation is used to control locoregional disease and chemotherapy is used to control micrometastatic disease
2. Nonoverlapping toxicity: chemotherapy and radiotherapy have different mechanisms of toxicity that are additive to the tumor but less potentiating to normal tissues.
3. Chemotherapy decreases the ability of radiotherapy-damaged tumor cells to repair.
4. Chemotherapy has cytotoxic activity against radioresistant (e.g., hypoxic) tumor cells.
5. Chemotherapy potentially has selective cytoprotective properties for normal tissues, allowing higher radiation doses (rationale for mucosal protectants).
6. Specific chemotherapeutic drugs such as fluorouracil, cisplatin, and taxanes induce cell cycle arrest at the G_2 checkpoint, where the cell is most radiosensitive.

Several different approaches have been used in concomitant chemoradiotherapy. The most frequently used has been standard radiotherapy or hyperfractionated radiotherapy with platinum-based chemotherapy. Phase III randomized trials and meta-analyses have shown statistically significant improved survival using concomitant chemotherapy and radiation for locally advanced squamous cell carcinomas for all sites in the head and neck. These trials have resulted in increased toxicity, including mucositis and systemic toxicity. These complications, however, have been manageable. Based on these trials, most major cancer centers now consider concomitant chemoradiotherapy a reasonable option for locally advanced carcinomas of the head and neck.

Nasopharyngeal carcinomas are distinct carcinomas of the head and neck. They occur in a younger age group, have a geographic distribution, metastasize early, and are more responsive to radiation and chemotherapy. There is a strong association between these tumors and Epstein-Barr virus, particularly in patients from southeastern China. The incidence of nasopharyngeal carcinoma in southeastern China has been reported as 10 to 20 per 100,000 men. Worldwide incidence of these tumors is less than 1 per 100,000, and it is a rare tumor in the United States.

Stage I and II lesions are treated with aggressive radiotherapy. Stage III and IV lesions are treated with chemoradiotherapy. Chemotherapy regimens usually have included cisplatin and fluorouracil with or without epirubicin. Five-year survival rates for stage I and II disease have been reported in a range of 50 to 90%, and for stage III and IV disease 17 to 60%. Unlike other carcinomas of the head and neck, surgery has a very limited role in the management of nasopharyngeal carcinomas.

Adjuvant Chemotherapy

Following definitive locoregional therapy, adjuvant chemotherapy has been given to control microscopic residual disease and micrometastatic disease. Adjuvant therapy has been proven to be of significant value in the management of breast and colorectal cancer. There have been a number of phase III trials of adjuvant therapy for head and neck cancers. Although there appears to be some reduction in the incidence of distant metastases, adjuvant therapy has not been demonstrated to improve survival. The vast majority of deaths related to head and neck cancer are due to local and regional failures and not to disseminated metastatic disease. At this time, adjuvant chemotherapy has not been shown to have a role in the management of carcinomas of the head and neck.

Palliative Chemotherapy

Chemotherapy has been used in the management of head and neck cancers that are recurrent, metastatic, unresectable, and considered incurable. It is in these settings that new systemic therapeutic modalities are initially evaluated in clinical trials. Phase I are dose-escalation trials evaluating toxicity and establishing maximum tolerated doses. They may involve single or multiple agents. Phase II trials are designed to determine the efficacy of new therapeutic regimens after their dose tolerance has been determined in phase I trials. Response rates are usually compared with previous trials (historic controls). Phase III trials usually are randomized and are always comparative. New therapeutic regimens are compared with standard therapy.

The primary goal of palliative therapy is to improve the patient's quality of life. This can be accomplished by relieving pain, preserving or improving organ function, and preventing obstruction of the airway or esophagus. Although in some instances survival may be prolonged, survival is not the primary goal of palliative therapy. Some clinical trials, however, have established prolonged survival with palliative therapy.

Single-agent methotrexate therapy has been the standard palliative therapy for head and neck cancer. It is well tolerated, convenient, and inexpensive. Response rates range from 15 to 30% with a median duration that generally has been less than 6 months. Multiagent chemotherapy with cisplatin and fluorouracil has demonstrated higher response rates (30–50%) and may prolong survival slightly. This regimen is more toxic, less convenient, and more expensive. It may be the regimen of choice for some patients with good performance status. Several new therapeutic regimens are being evaluated for recurrent and metastatic squamous cell cancer. These include taxanes, vinorelbine, and gemcitabine in single- and multiagent trials.

Quality of Life

HNSCC is a devastating illness. Following treatment, patients often experience problems with speech, swallowing, and disfigurement. Although multiple treatment approaches are often available for a given HNSCC patient, not all treatments will have the same impact on the patient's quality of life. Choice of treatment will often result in a trade-off between chance of cure and the patient's quality of life. There are good data regarding cure rates associated with various treatment regimens. We are starting to see good data comparing these regimens (e.g., radiation alone vs combined chemoradiotherapy). Unfortunately, data on quality of life are lacking. Quality of life outcomes research relies on patients' reports. Data should be collected prospectively and include validated measures of social, emotional, physical, and psychological function. This field of HNSCC research is in the early phases of development, but it will ultimately provide important information to patients and oncologists who together face difficult treatment decisions.

MOLECULAR BIOLOGY OF HEAD AND NECK CANCER

Advances in our understanding of HNSCC molecular biology give promise to the future development of more sophisticated methods to prevent, diagnose, and treat this cancer. Through these efforts significant reductions in morbidity and mortality are envisioned.

There is ample evidence that UADT mucosa undergoes a stepwise progression to squamous cell carcinoma. This is seen at the clinical, histological, and molecular levels. At the clinical level, affected mucosa progresses from leukoplakia or erythroplasia to raised or ulcerated lesions harboring squamous cell carcinoma. At the histological level, this progression can be seen as worsening dysplasia that develops into carcinoma in situ and finally invasive carcinoma on serial biopsies. On the molecular level, tumor tissues show escalating numbers of genetic alterations. In most cases the accumulation of genetic alterations is a result of exposure to tobacco and alcohol. Other risk factors, such as viral exposure and heredity, also play a role, as outlined previously. Although the exact order and number of genetic events required for cancer development are still under investigation, six important steps are believed to be necessary. These are loss of growth inhibition signals and programmed cell death; immortalization; and acquisition of autonomous proliferation, angiogenesis, and the ability to invade tissue.

Oncogenes are genes whose protein products are involved in cell growth regulation so that overexpression or mutation results in disregulated cell growth. Epidermal growth factor receptor (EGFR), a member of the erbB family of receptor tyrosine kinases, is one such oncogene that has been studied in HNSCC. EGFR and its ligand, transforming growth factor α (TGF-α), are overexpressed in 90% of HNSCC. Binding of TGF-α, EGFR, and other growth factors to the extracellular domain results in phosphorylation and receptor activation that stimulates tumor growth through autocrine and paracrine pathways. Oncogenes overexpressed in HNSCC include other growth factors or growth factor receptors (hst1/fibroblast growth factor 4, int2/fibroblast growth factor 3, Her2/erbB2), intracellular signal transducers (ras, stat3), transcription factors (c-myc/N-myc), regulators of cell cycle (cyclin D_1), and inhibitors of apoptosis or programmed cell death (Bcl2).

Tumor suppressor genes inhibit tumor cell growth. Such genes are involved in cell cycle arrest and apoptosis. The p53 is the most studied tumor suppressor gene in tissue specimens of head and neck cancer. It is located on chromosome 17p13 and encodes a nuclear protein that regulates the G1/S transition of the cell cycle (see Chapter 19). It can induce cell-cycle arrest and DNA repair or apoptosis in response to genetic damage. Inactivation of p53 most often occurs by point mutation, although rearrangements, deletions, and inactivation by viral proteins are also possible. Studies have reported mutations in 25 to 69% and overexpression in 15 to 60% of oral cancers. Other important tumor suppressor genes that are frequently inactivated in HNSCC include inhibitors of cell cycle progression (p16, p21, and $p27^{kip}$).

Telomerase is an enzyme thought to be involved in cell immortalization; telomeres are highly conserved, repeated DNA sequences located at the ends of chromosomes. Normally, cells fail to replicate the 5′ end of linear DNA, resulting in progressive loss of telemetric DNA with each cell division. It is believed that this loss of DNA signals the cell to enter senescence.

The telomerase enzyme adds DNA sequences before each cell division, resulting in immortalization of the cell. Telomerase is usually only active in germ cells and hematopoetic cells, but is present in most cancers, including HNSCC.

Matrix metalloproteinases (MMPs) are important regulators of the extracellular environment. MMPs and related proteins (A Disintegrin And Metalloproteinase [ADAMs]) represent a family of proteinases that can process almost any component of the extracellular matrix. These proteinases are down-regulated by inhibitors such as tissue inhibitors of metalloproteinases (TIMPs). MMPs process molecules involved in cell–cell and cell–matrix interactions such as CD44 and integrin molecules. Disregulation of MMPs and related proteins can result in cell dissociation and invasion. MMPs also cleave growth factor–binding proteins, releasing mitogens that lead to cell division and inhibition of apoptosis. Some MMPs, including MMP1 and MMP3, have been shown to be overexpressed in HNSCC and associated with lymph node metastases.

Cyclooxygenase is an important regulatory enzyme in the production of prostaglandins from arachidonic acid. Cyclooxygenase-1 is produced under normal conditions. Cyclooxygenase-2 (COX-2) is induced by pathophysiological factors such as inflammatory stimuli and oncogenes and is overexpressed in premalignant and malignant UADT lesions. Prostaglandins produced by COX-2 exert their effect by binding to both cell surface and nuclear hormone receptors and are involved in cell proliferation, angiogenesis, antiapoptosis, invasion, and metastases.

Retinoids are derivatives of the natural compound retinal (vitamin A). Retinoids mediate gene trascription by binding to retinoic acid receptors. Retinoic acid receptors comprise a family of homo- and heterodimers. When activated, they bind to the promoter region of genes, resulting in antiproliferative and proapoptotic effects.

Chemoprevention

Chemoprevention includes strategies to prevent or reverse carcinogenesis before an invasive cancer develops or to prevent a second primary cancer in patients who have had a previous cancer cured.

Clinical trials with various retinoids have studied histopathologically confirmed oral premalignant epithelial dysplastic lesions. Hong et al used high-dose isotretinoin (13-*cis*-retinoic acid, Accutane) in 44 patients with dysplastic oral epithelial lesions; 67% of the patients had major responses, compared with a 10% response in the placebo group. Within 3 months after discontinuing therapy, 50% of the lesions recurred. High-dose isotretinoin was associated with significant side effects, including cheilitis, conjunctivitis, and hypertriglyceridemia.

In a randomized study using high-dose isotretinoin induction followed by maintenance with either low-dose isotretinoin or beta-carotene, 8% of the patients on isotretinoin had progression of their oral lesions; 55% of the beta-carotene group had progression of oral lesions. Maintenance therapy was not associated with significant side effects. Other studies have evaluated vitamin A, retinamide, and fenretinide, and have not proven them to reverse oral premalignant lesions.

Several phase three studies have been conducted using retinoids for the prevention of second primary cancers in patients who had received definitive therapy for a primary head and neck cancer. Initial reports showed reduction of second primary tumors with a high-dose regimen. However, follow-up studies failed to show adequate benefit.

Retinoids have significant toxicity in the high doses used in these studies Molecular biology has provided new information on how retinoids regulate gene expression and has led to the development of synthetic retinoids that may be less toxic and more effective in the prevention of cancer.

A study comparing beta-carotene with placebo over 12 years in 22,000 physicians showed essentially no difference in the incidence of cancer, stroke, myocardial infarction, and death. A study comparing beta-carotene plus vitamin A with placebo was stopped after 4 years because of a significantly increased incidence of lung cancer and overall mortality in the study group. COX-2 inhibitors currently are under investigation in phase II clinical trials as chemopreventive agents in patients with leukoplakia and dysplasia. At this time, chemopreventive therapy for head and neck cancer must be considered investigational. The use of chemopreventives should be limited to controlled clinical trials.

Early Detection

No symptom or symptom complex is strongly correlated with HNSCC for any subsite except the larynx; consequently, most HNSCC patients present at late stage. Because oral cavity cancer is the most common site for HNSCC and it is fairly easy to examine, it seems that screening examinations would be beneficial. However, a systematic review concluded there is insufficient evidence to assess the effectiveness of community-based screening programs. Despite the American Cancer Society's and National Cancer Institute's emphasis that oral examinations could prevent many deaths, only 16% of respondents to the Centers for Disease Control and Prevention's 1998 National Health Interview Survey had ever undergone an oral cavity exam. Selected high-risk populations are most likely to benefit from screening exams. Detection rates have reached 2.4% in highly targeted populations compared with no cases to 0.1% for general screening.

Adjunctive tools to the physical exam include toluidine blue vital dye used as a rinse or applied directly to suspicious lesions. Toluidine blue has yielded specificities between 45 and 93% and sensitivities between 72 and 100%. The oral cytobrush that generates a cytology specimen has yielded impressive results in the hands of academicians. There is insufficient evidence to evaluate the effectiveness of toluidine blue and oral cytobrush in high-risk community settings.

Current research focuses on molecular approaches for accurate prognosis and early detection. Although epithelial dysplasia remains the most important predictive factor, DNA content, aberrant oncogenes, integrin expression, p53 expression, and loss of heterozygosity at chromosome 3p or 9p show some promise in predicting which lesions will progress to cancer. Loss of heterozygosity refers to loss of one allele of a gene. Loss of one allele of a tumor suppressor gene through chromosomal deletion and the other allele through a different mechanism will lead to a functional deficit of the tumor suppressor. Microsatellites are repetitive DNA sequences that are scattered throughout the genome. Microsatellite instability refers to the expansions and deletions within these regions that are frequently seen in various cancers. These differences in macrosatellites between normal and tumor samples can be identified by polymerase chain reaction (PCR) and used to detect regions of allelic loss. DNA promoter hypermethylation also leads to inactivation of tumor suppressor genes and can be detected by PCR methods. Hu and colleagues (2002) reviewed molecular detection approaches for smoking associated tumors. Several markers have been studied in HNSCC saliva and serum. Spafford and colleagues showed loss of heterozygosity or microsatellite instability in 1 of 23 markers in 79% of saliva samples from 44 HNSCC patients and in none of 43 healthy control subjects. Boyle et al identified tumor-specific p53 mutations in five of seven saliva samples from patients with HNSCC. Abnormal promoter hypermethylation has been detected in the saliva of 11 of 17 HNSCC patients. Mitochondrial DNA mutations have been detected in six of nine saliva samples from HNSCC by direct sequencing. Franzman's group showed that hyaluronic acid was elevated in eight and hyaluronidase elevated in saliva of 11 of 11 HNSCC patients. Microsatellite markers have been identified in the serum of 6 of 21 patients with HNSCC. In 50 patients with paired serum and tissue, the same promoter hypermethylation pattern in one of three genes (p16, MGMT, and DAP-K) was found in 42% of patients.

High throughput techniques have revolutionized the search for head and neck tumor markers. Ribonucleic acid (RNA) or DNA samples are hybridized to complementary (c)DNA arrays that permit analysis of thousands of genes simultaneously. Similarly, surface-enhanced laser desorption/ionization–time of flight (SELDI-TOF) mass spectrometry and antibody microarrays have been used to analyze samples from HNSCC patients and normals to develop a protein-level signature for HNSCC. Although these methods may not be feasible for mass screenings, they facilitate identification of the most important markers. Once reliable markers are identified, less expensive screening modalities such as enzyme-linked immunosorbent assay (ELISA) can be developed for early-detection programs.

Predicting Response and Prognosis in Head and Neck Carcinoma

Tumors exhibit wide variations in their response to radiation and chemotherapy. Recent research has focused on the identification of markers that may predict the response of an individual tumor to cytotoxic radiation and/or chemotherapy. A number of biological markers have been shown to be predicative of poor response to radiation and/or chemotherapy. Identified markers include p53 mutations, p53 overexpression, epithelial growth factor receptor ($erbB_1$) overexpression, $erbB_2$ mutations, cyclin D_1 (BCL1) overexpression, tumor hypoxia, angiogenic factors, tumor microvascular density, and thymidylate synthetase.

Head and neck carcinomas that express high levels of thymidylate synthetase have been shown to be more resistant to fluorouracil therapy than tumors with low levels of thymidylate synthetase. In view of this, chemotherapy for tumors with high levels of thymidylate synthetase should include agents other then fluorouracil.

Previously, tissue electrodes have measured tissue oxygenation. Recently, positron emission tomography, which can more accurately identify and map tissue oxygen levels, has become available. This may allow for prospective prediction of appropriate therapeutic agents for individual tumors. Squamous cell carcinomas of the head and neck have a disorderly microvascular supply that results in areas containing large numbers of hypoxic cells. These hypoxic cells are resistant to radiation and to certain chemotherapeutic drugs. Conversely, bioreductive drugs such as mitomycin C have increased cytotoxicity in areas of hypoxia. The cytotoxicity of drugs such as fluorouracil and cisplatin are not affected by levels of tissue oxygenation.

Hypoxic areas in tumors induce the formation of angiogenic factors such as vascular endothelial growth factor (VEGF). These angiogenic factors result in areas of increased tumor vascularity, which have increased resistance to some chemotherapeutic agents. Tumor microvascular density, which can be induced by angiogenic factors, is a known predictor of poor tumor response. Tumors with high microvascular density may respond to antiangiogenic factors. The predictive value of these biological markers needs to be validated in prospective therapeutic trials. If validated biological markers may become of great value in customizing therapy for specific carcinomas.

Clinical Trials

Several strategies have been used to inhibit VEGF-mediated signals. These include anti-VEGF antibodies and agents that inhibit the VEGF receptor tyrosine kinase. Phase III trials using anti-VEGF agents in various cancer types, in combination with chemotherapeutic agents, are ongoing.

Gene therapy is currently being evaluated in several phase II clinical trials. One trial involves transfection of the normal p53 gene into carcinomas with mutant p53 using an adenovirus vector. Another phase II trial used a mutant adenovirus that replicates selectively in p53-deficient human tumor cells, resulting in tumor cell lysis, and objective tumor responses were demonstrated. The adenovirus does not replicate in normal cells with functional p53 genes.

Therapeutic approaches targeting EGFR are also under investigation. Strategies include blockade of the extracellular receptor domain and inhibition of the tyrosine kinase activity, among others. Most of these studies are in phase I–II testing. EGFR antagonists are being studied alone (with modest effects reported) as well as in combination with chemotherapy or radiation. Results of these studies are pending.

SUMMARY

In recent years, we have seen several advances in the management of head and neck cancer. Chemotherapy has been shown to be of value in organ preservation, thus decreasing morbidity. Present data support the use of concomitant chemotherapy and radiation therapy as treatment for locally advanced squamous cell carcinoma of the head and neck where surgery is not a good option. Advances in microvascular surgery have resulted in improved reconstruction and function following surgical extirpation of these cancers. Future advances in immunology and molecular biology promise to provide us with greater knowledge about the etiology, diagnosis, prognosis, and prevention of cancer. With this new knowledge, utilizing immunotherapy and gene therapy, along with improved traditional head and neck cancer treatment, cancer will be more effectively managed with less morbidity and better survival. New knowledge about individual cancers should also allow for more customized therapeutics based on biological markers that predict more accurately an individual tumor's behavior and response to specific therapeutic modalities.

Suggested Readings

Boyle JO, Mao L, Brennan JA, et al. Gene mutations in saliva as molecular markers for head and neck squamous cell carcinomas. Am J Surg 1994;168:429–432

Cancer Facts and Figures—2003. Atlanta: American Cancer Society, 2003.

Chung CH, Parker JS, Karaca G, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 2004;5:489–500

Das B, Nagpal J. Understanding the biology of oral cancer. Med Sci Monit 2002;8:258–267

Ferlito A, Silver C, Howard D, et al. The role of partial laryngeal resection in current management of laryngeal cancer: a collective review. Acta Otolaryngol 2000;120:456–465

Franzmann EJ, Schroeder GL, Goodwin WJ, et al. Expression of tumor markers hyaluronic acid and hyaluronidas (HYAL1) in head and neck tumors. Int J. Cancer 2003;106:438–445

Forastiere AA, Goepfert H, Maor M, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. N Engl J Med 2003;349:2091–2098

Forestiere A, Koch W, Trotti A, Sydransky D. Head and neck cancer. N Engl J Med 2001;345:1890–1898

Fu KK. Combined radiotherapy and chemotherapy for nasopharyngeal carcinoma. Semin Radiat Oncol 1998;8:247–253

Fu KK, Pajak TF, Trotti A, et al. A Radiation Therapy Oncology Group (RTOG) phase III randomized study to compare hyperfactionation and two variants of accelerated fractionation radiotherapy for head and neck squamous cell carcinomas: first report of RTOG 9003. Int J Radiat Oncol Biol Phys 2000;48:7–16

Ha P and Califano J. The molecular biology of laryngeal cancer. Otolaryngol Clin N Am 2002;35:993–1012

Hong W, Endicott J, Itri LM, et al. 13 Cis-retinoic acid in the treatment of oral leukoplakia. N Engl J Med. 1986;315:1501–1505

Hu Y, Sidransky D, Ahrendt S. Molecular detection approaches for smoking associated tumors. Oncogene 2002;21:7289–7297

Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. The Department of Veterans Affairs Laryngeal Cancer Study Group. N Engl J Med. 1991;324:1685–1690

Kim E and Hong W. An apple a day . . . Does it really keep the doctor away? The current state of cancer chemoprevention. J Natl Cancer Inst 2005;97:468–469

Kirchner J and Carter D. Intralarygeal barriers to the spread of cancer. Acta Otolaryngol (Stockh) 1987;103:503–513

Mohan S and Epstein J. Carcinogenesis and cyclooxygenase: the potential role of COX-2 inhibition in upper aerodigestive tract cancer. Oral Oncology 2003;39:537–546

Morton R and Izzard M. Quality-of-life outcomes in head and neck cancer patients. World J Surg 2003;27:884–889

Napgal J, Das B. Oral cancer: reviewing the present understanding of its molecular mechanism and exploring the future directions of its effective management. Oral Oncol 2003;39:213–221

National Cancer Institute, National Institutes of Health, Public Health Service, Department of Health and Human Services. Oral cancers: research report. (NIH publication no. 92-2876). Bethesda, MD;1991

Parsons J, Mendenhall W, Stringer S, et al. Twice-a-day radiotherapy for squamous cell carcinoma of the head and neck: the University of Florida experience. Head Neck 1993;15:87–96

Spafford MF, Koch WM, Reed AL, et al. Detection of head and neck squamous cell carcinoma among exfoliated oral mucosal cells by microsatellite analysis. Clin Cancer Res 2001;7:607–612

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. A patient with a right 3.5 cm oral tongue squamous cell carcinoma with normal tongue mobility, no mandibular involvement, no distant metastases, and two 2 cm lymph nodes on the right side is classified as
   A. T2N1MO, stage II
   B. T3N2bMO, stage III
   C. T2N2aMO, stage III
   D. T2N2bMO, stage IV
   E. T2N2bM0, stage III
2. Squamous cell dysplasia
   A. is a clinical diagnosis
   B. is equivalent to leukoplakia
   C. transforms to invasive squamous cell carcinoma 36 to 61% of the time
   D. is one of the best known predictors of invasive squamous cell carcinoma
   E. most often occurs after the sixth decade of life
3. In an otherwise healthy patient with advanced laryngeal cancer, the treatment regimen that is most likely to result in preservation of the larynx at 2 years is
   A. concurrent chemotherapy and radiation therapy
   B. induction chemotherapy followed by radiation.
   C. radiation followed by adjuvant chemotherapy
   D. total laryngectomy followed by radiation
   E. radiation alone
4. Examples of tumor suppressor genes include all of the following except
   A. p53
   B. p16
   C. cyclin D_1
   D. erbB2
   E. c and d
5. COX-2
   A. is involved in prostaglandin synthesis
   B. is usually active under normal conditions
   C. is not expressed in head and neck squamous cell carcinoma
   D. all of the above
   E. a and c

Chapter 10B
IMMUNOBIOLOGY AND IMMUNOTHERAPY OF HEAD AND NECK SQUAMOUS CARCINOMA

GIOVANA THOMAS, WILLIAM J. RICHTSMEIER, AND HARI NADIMINTI

HISTORICAL PERSPECTIVE

The immunology of head and neck squamous cell carcinoma (HNSCC) of the upper aerodigestive tract occupies a special place in tumor biology. It is therefore important to define what cancers are being discussed. The term *HNSCC* usually excludes squamous cell carcinomas of the skin, which seem to have a somewhat different biology from mucosal-derived cancers even though the head and neck surgeon usually will provide care for both. Similarly, the biology of cancers that arise from the paranasal sinuses, nose, and nasopharynx have a different induction from other upper aerodigestive tract carcinomas. Other tumors that appear to have a separate biology are melanomas that occur in the head and neck and glandular-based (adeno-) carcinomas. These tumors have different patterns of induction, spread, and response to therapy. Although there are some similar fundamental concepts shared in all cancers, this chapter will focus on the immunobiology of squamous cell carcinoma of the upper aerodigestive tract arising in the oral cavity, oropharynx, hypopharynx, and larynx.

The identification of normal cells and tissues by the body is a complex and complicated phenomenon. For years, it was thought that a unique activity occurred during the newborn period in mammals that allowed maturation of the immune system, whereby a cell learns to distinguish self from nonself. It was known that identical twins of various species could receive transplants from one to another if the immune exposure (i.e., shared circulation via the placenta) occurred during gestation or in the early newborn period. Once the observation that dizygotic cattle twins who share common placentas could receive organ transplants from one another (even though they were separate genetic individuals), the advance of transplantation biology leaped ahead. By this time, the concept of surface antigens of major blood groups had been identified. Whereas under normal circumstances individuals with major blood group incompatibilities would immediately reject organs from one another, dizygotic twin cattle would accept organ transplants later in life from each other, but not from siblings from other pregnancies.

A series of experiments attributed to Medawar and colleagues investigated transplanted hematologic stem cells from one group of mice to another. Newborn mice of one species who were given stem cells from a second species would later on in life accept skin transplants from the second species, whereas the control animals not receiving stem cells would reject them in the normal way. These experiments from the 1950s became the basis for the concepts of immunologic development that would predominate for the next 45 years.

Over the past several years, the concept of antigen-processing cell control of these activities has become more important. The term *professional antigen-processing cells* specifically refers to dendritic cells and a few other cells that are largely responsible for these phenomena. Matzinger and others have been able to show that both tolerance and immunity can be induced in the newborn period as well as later in life. Matzinger's experiments use inbred mice that are immunologically identical from one individual to the other. Mice, however, are different from humans, in that male mice express a unique antigen, Hy, on the surface of all cells. Male mice have one more antigen than female mice. This difference allows experiments in which all but one antigen can be controlled in an experiment. What Matzinger observed was that spleen cells from male mice given to female mice in the newborn period would induce tolerance similar to the experiments observed by Medawar et al. If, however, dendritic cells from male mice were purified and used as the material to be injected, specific anti-male cell-mediated immunity was induced. This immunity could be induced either in the newborn period or in immunologically mature life; in a corresponding fashion, large numbers of hemopoietic cells that contained few, if any, dendritic cells and, even given later in life, induced tolerance similar to that which was observed in the newborn period.

Conclusions from these experiments are that the dendritic cells are responsible for inducing immunity either in the newborn period or later in life. The explanation of Medawar et al's experiments is that the inocula of "stem cells" into the newborn mice included mostly B cells and T cells but did not contain a critical number of professional antigen-processing cells, which are the only cells that can activate a virgin T cell. Because the inocula of "stem cells" contains large numbers of immunologically important cells that can induce tolerance such as B cells and cells that are immunologically neutral, such as non-naive T cells, the end result is to induce tolerance.

The interaction between dendritic cells and T cells has been studied in great detail, and although there is still a great deal left to be uncovered, it appears that dendritic cells must present the antigen in combination with a major histocompatibility antigen complex. This presentation is not a selective process, and many antigens can be expressed in this situation, which are pinocytosed by dendritic cells or by chance manufactured within the dendritic cells themselves. For the T cell to become activated, the dendritic cell must be in an activated state. Dendritic cells expressing normal antigens on their surface when not in an activated state—induced tolerance.

Dendritic cell activation primarily occurs in what is commonly known as "infection." At least one mediator of this danger signal appears to be heat shock protein, a substance liberated early in the infectious process. Early cancers have no reason to call attention to themselves and initiate an infectious response. It is known that some tumors can produce substances rendering the dendritic cell inactive. Hence dendritic cells presenting tumor antigens on their surface are seen as presenting antigens to which the host should not respond; and when a dendritic cell presents an antigen in a nonactivated state, tolerance is induced. There is speculation that thymic T cells can never respond to the second signal; therefore, the process of thymic maturation is a statistical game of the likelihood that an antigen will be exposed to a nonreacting dendritic/T-cell complex, thus inducing tolerance.

The observations of the 1960s that demonstrated profound systemic and regional immunosuppression in head and neck cancer patients point to more than subtle disturbance in the immune system. It is likely that there are multiple ways in which the immune system is affected by cancer and that one single therapy affecting the immune system will not completely reactivate an immune system that has become tolerant in a patient with advanced cancer. Future experiments will need to involve patients with relatively early tumors who still have intact immune systems that can be activated, with the hope that, once the immunologic mechanisms are understood, immunologic manipulation can be used to treat postsurgical patients.

IMMUNOBIOLOGY OF HEAD AND NECK SQUAMOUS CELL CARCINOMA

HNSCC of the upper aerodigestive tract is a devastating disease that affects both communication and survival. HNSCC represents ~4% of all cancers, with an estimated

37,000 new cases and 12,000 expected deaths in the United States in 2003. Approximately half of all patients have advanced-stage disease at the time of diagnosis, with an expected 5-year survival rate between 10 and 40%. Despite treatments that may consist of mutilating surgery, radiotherapy, and/or chemotherapy, overall long-term survival remains low due to uncontrollable persistent or recurrent disease. The low rate of survival of patients with local and distant recurrence has highlighted the need for new approaches for diagnosis and treatment. Stimulating the immune system to recognize and destroy tumor cells is a promising treatment modality for HNSCC patients, who have been shown to have significant immunologic dysfunction. However, the intricate cross-talk between cancer and cells of the immune system is still poorly understood in HNSCC.

Recent breakthroughs in the field of tumor immunology have provided tools to better understand the complex interactions between squamous cancer cells and their tumor environment. Dendritic cells have been recognized as one of the key players necessary for induction of immune responses to cancer and immunosurveillance. Genes encoding tumor-associated antigens (TAAs) have been identified, and strategies for immunizing against these antigens have been investigated. Indeed, immunotherapeutic strategies have shown that immune manipulation can induce the regression of established cancer in humans. This chapter will summarize our knowledge of tumor immunology in HNSCC and will discuss recent advances in immunotherapy that hold promise for eliciting and sustaining antitumor immune responses in these patients.

Deficient Immune Responses in HNSCC

Like many cancers, the interrelations between transformed epithelium and the tissue microenvironment play a critical role in tumorigenesis of squamous cell carcinoma (SCC). In particular, experimental and clinical evidence supports the concept that deficiencies in the immune system play an important role in the initiation and propagation of HNSCC. First, patients who are immunosuppressed due to heart or kidney transplant treatments show increased frequencies of SCC of the skin. Moreover, precancerous lesions in transplant patients progress more rapidly into SCC and show an increased tendency to metastasize, suggesting a failure of immunosurveillance. Second, increased antitumor immune responses in patients with HNSCC have been demonstrated in draining lymph nodes and tissues harboring HNSCC after treatment with immunostimulatory agents such as bacille Calmette-Guérin or levamisole, suggesting that it is feasible to elicit immune responses in these patients. Third, the presence of tumor-infiltrating lymphocytes (TILs) in tumor specimens of patients with HNSCC has been shown to collaborate with better clinical outcomes in these patients, suggesting that TILs may play a significant role in HNSCC tumor regression. This evidence suggests that the immune system plays an important role in surveillance and tumor growth of HNSCC, and also suggests that growth of these tumors may be prevented or modified by interventions that target immune responses. Further basic scientific research is required before we completely understand how HNSCC evades immune elimination and use this information to devise successful methodologies to increase immune killing.

Failure of Immunosurveillance in HNSCC

The role of immunosurveillance is to protect the host from tumor formation. In order for the immune system to accomplish this task, several criteria must be met. First, the host should have a competent immune system with a strong cytotoxic response. Second, the abnormal/tumor cells must express unique antigens that can be recognized by the immune system. Third, there should be no suppressive influences to dampen the immune response. Lastly, the number of abnormal/tumor cells should be small enough for the tumor to be eliminated entirely.

Several mechanisms may account for failure to elicit an effective antitumor immune response against developing HNSCC.

Malignant transformation of epithelial cells may involve loss or changes of normal surface antigens. Downregulation of immune recognition antigens on the surface of tumor cells is a potential mechanism used by HNSCC to escape from immunosurveillance and killing by cytotoxic T cells. Changes in the immune recognition antigens major histocompatibility complex (MHC) class I [human leukocyte antigens, (HLAs), in humans], and/or costimulatory molecules may block recognition and allow for tolerance of the immune system to tumor cells. A majority of tumors from patients with HNSCC show no expressions of HLA-class I, and a small percentage of these tumors show extensive downregulation of this molecule. Inadequate expression of costimulatory molecules on the surface of tumor cells may also contribute to immune tolerance. Costimulatory molecules are important for T-cell activation. CD80 and CD86 molecules provide the most potent costimulatory signals found to date. Many cell lines and established HNSCC lack expression of costimulatory molecules on the surface of tumor cells. In addition, decreased or lack of CD80 costimulatory molecule expression has been associated with increased tumorigenesis of HNSCC.

The absence of an effective cell-mediated antitumor immune response may be due to poor immunogenicity of TAAs. These are antigens expressed by tumors that distinguish them from normal cells. Although TAAs are found on tumor cells and also on some normal cells, qualitative and quantitative differences in their antigen expression may permit the immune system to distinguish between the two. Integrins, mucins, cadherins, growth factor receptors, and glycoproteins are types of TAAs overexpressed on the cell surface that may play an important role in carcinogenesis and progression of HNSCC. Despite their expression, these antigens may evoke an immune response that is successfully evaded or no response at all. Methods to elicit therapeutic responses to these antigens in patients whose tumors express the protein are being investigated.

Soluble factors secreted during the progressive growth of HNSCC have been shown to restrict cell-mediated immune functions resulting in immunosuppression of the host. Autocrine cytokines secreted by HNSCC such as granulocyte-macrophage colony–stimulating factor (GM-CSF) have been shown to have direct immunosuppressive effects on host immune responses. In addition, cytokines secreted by precancerous or HNSCC lesions may lead to increased expression of COX-2 enzyme in tumor-associated macrophages and, subsequently, increased synthesis of prostaglandins. Prostaglandins have been shown to decrease T-cell proliferation and natural killer (NK) cell cytotoxicity and to inhibit the production of immune regulatory lymphokines.

The Role of Cell-Mediated Immunity in HNSCC

Tumor-specific immune responses have been demonstrated during the growth of established HNSCC, despite the failure of the immune system in preventing its emergence. Immune responses to these tumors, however, generally are ineffective in inducing regression of established HNSCC. Several immunologic abnormalities in cell-mediated immune responses in patients with HNSCC have been observed. These include decreased T-cell proliferation, migration, and cytotoxicity responses, depressed NK cell activity, significant dysfunction of T helper immunity, and depressed lymphokine-activated killer cell activity.

The T-cell immune response is the most important host response for the control of growth of HNSCC. T-cell immunity involves two T-cell subsets: CD4+ T helper cells, which are class II MHC restricted, and CD8+ T cytolytic cells, which are class I MHC restricted. CD4+ T cells mediate their effect by the generation of lymphokines/cytokines to activate other effector cells. CD8+ T cells also can secrete cytokines, but they are involved in direct lysis of tumor cells by disrupting the target membrane and nucleus. Perturbations in the numbers and function of T lymphocyte subsets in patients with HNSCC have been associated with decreased survival. Recent evidence suggests that a significant proportion of circulating T cells (especially CD8+ T cells) in patients with HNSCC is eliminated by apoptosis (programmed cell death). The reasons for this are still not fully understood. However, inadequate stimulation of T cells by lack of CD80 costimulatory pathway signals on antigen-presenting cells (APCs) at the time of T-cell stimulation is a possible mechanism that may cause apoptosis of T cells.

Dendritic cells are important in the induction of immunity against HNSCC. They are the most potent of the APCs, whose function is to process and present antigens to T cells in the context of MHC. T cells must receive two signals to become activated. One signal is mediated by the interaction between T cell and peptide bound to MHC molecules on the surface of APCs. The second signal is provided by costimulatory molecules on the APCs interacting with their receptors on the surface of T cells. Increased infiltration of dendritic cells into tumors has been shown to correlate with tissue differentiation, prolonged survival rates, and reduced incidence of metastasis in patients with HNSCC of the nasopharynx, oral cavity, and larynx. Furthermore, tumor dendritic cells in patients with HNSCC have been shown to lack adequate antigen-presenting function. These studies suggest that T cell–mediated immunity to HNSCC may depend on the satisfactory activation of dendritic cells.

NK cells are large lymphocytes that play a critical role in the early host defense against cancer. In contrast to T cells, NK cells do not require prior sensitization to antigen to effect killing of tumor cells. NK cells produce large amounts of immunoregulatory cytokines in response to various activation stimuli and can therefore regulate T-cell immune responses. Suppression of NK cell function has been observed in draining lymph nodes of patients with HNSCC. Deficient NK cell activity has been correlated to nodal metastases in patients with laryngeal carcinoma. Thus low pretreatment NK cell activity in peripheral blood of patients with HNSCC may serve as a prognostic indicator for survival. In patients with poorly differentiated, low HLA class I–expressing molecules, the NK cell is an important defense mechanism against metastatic disease.

In summary, host–immune system interactions in HNSCC comprise an intricate molecular network still poorly understood. Immunosuppression and tumor progression in these patients may be the net result of a complex cascade of events that begins with failure of immunosurveillance. Successful immunotherapeutic approaches in the treatment of HNSCC will need to repair host immunoincompetence in the function of T and NK cells, circumvent immunosuppressive factors generated in the tumor microenvironment, optimize target tumor antigen presentation, and sustain a persistent long-term effective antitumor immune response.

IMMUNOTHERAPY OF HNSCC

The major goal of immunotherapy of cancer is to manipulate and amplify the immune system to promote tumor eradication. As the field of tumor immunology of HNSCC advances and pathways of immune dysfunction in HNSCC are delineated, considerable emphasis is being placed on identifying specific immunologic abnormalities that may prove to be useful prognostic markers and targets for immune therapy. The objectives to achieve successful immunotherapy in HNSCC are to activate local defense factors, induce lymphocyte release of mediators such as cytokines, and increase direct lymphocyte and macrophage killing of tumors.

The strategies employed by immunotherapy can be classified as (1) altering cytokine production, (2) optimizing the presentation of antigens to the immune system, and (3) transferring immune cells.

Altering Cytokine Production

Cytokines are a diverse group of intercellular signaling proteins that regulate local and systemic immune responses. When cytokines are secreted by tumor cells or by tumor-infiltrating lymphocytes, they appear to regulate tumor cell growth and the cytotoxicity of TILs. However, patients with HNSCC have a deficit of immune stimulatory cytokines present in the tumor milieu. Increasing immune stimulatory cytokine levels has been a potential target because it is thought to increase the local production and activation of CD8+T cells and NK cells. Interleukin-2 (IL-2), IL-12, and interferon γ are polypeptides that have been studied in this context. Amplification of cell-mediated immune responses requires an optimal supply of these cytokines.

IL-2 is a polypeptide secreted by CD4+ T cells that induces lymphokine-activated killer (LAK) activity and activates neutrophils and macrophages. As a result, it enhances the immune responses by increasing production of secondary cytokines. Despite achieving modest success in the treatment of renal cell carcinoma and metastatic melanoma (15–20% antitumor response rate), IL-2 by itself has not been particularly effective in inducing significant tumor regression in HNSCC. Initial studies using IL-2 therapy in the late 1980s and early 1990s showed disappointing results with minimal (6–10%) partial response rates observed that lasted less than 6 months. Toxicity from systemic administration of IL-2 also limits its clinical use. However, new studies show some promise using IL-2 (intratumoral or perinodal) in combination therapy with other known chemotherapeutic agents or cytokines. Potentially, this will allow for reductions in IL-2 dosages and, consequently, a decrease in systemic toxicities. Combination of external beam radiation or chemotherapy (e.g., cisplatin) with IL-2 has shown potent antitumor effects in preclinical models. When cisplatin is used in combination with IL-2, an increase in apoptosis and decrease in tumor growth are observed and may be due to IL-2's effect on activating CD8+ T cells and NK cells. IL-2 in combination with IL-12 has been shown to have synergistic effects on tumor control. Although these preclinical studies are encouraging, a phase III study failed to show a difference in the addition of IL-2 to cisplatin and 5-Fluorouracil (5-Fu) therapy in patients with HNSCC stages III and IV. Clinical trials using IL-2 with a combination of other cytokines such as interferon and tumor necrosis factor in patients with stage II HNSCC have resulted in clinical and histological tumor responses and appear promising, although these trials are still in the early phases of testing.

IL-12 is a protein produced by B cells, monocytes, and macrophages that augments the cytolytic activity of NK cells, synergizes with IL-2 to generate LAK cells, and facilitates the development of cellular immune responses. In preclinical models, use of IL-12 has shown significant antitumor responses and complete tumor regression. However, earlier clinical trials using systemic IL-12 in patients with advanced, unresectable solid tumors including HNSCC has been limited by its toxicity. A recent phase I clinical trial effectively limited toxicity of this cytokine by using peritumoral IL-12 gene transfer (IL-12-transduced autologous fibroblasts). This method has made IL-12 treatment feasible and has paved the way for future clinical studies to evaluate its potential.

Interferon γ is a polypeptide secreted by CD8+ T cells, some CD4+ cells, and NK cells. It increases surface expression of MHC molecules and activates many immune cells, including NK cells, neutrophils, vascular endothelial cells, and macrophages to secrete cytokines. Small-scale studies using interferon γ in patients with HNSCC showed disappointingly minimal clinical responses to systemic therapy. Further preclinical studies and large-scale clinical trials are required to determine whether increasing levels of cytokines IL-2, IL-12, and interferon γ have significant effects on tumor responses and clinical outcome.

In contrast to the cytokines already mentioned, some proinflammatory and proangiogenic cytokines such as IL-6, IL-8, and granulocyte-macrophage colony–stimulating factor (GM-CSF) are expressed at high levels by HNSCC and have been associated with potentially deleterious effects, including increased proliferation, angiogenesis, metastasis, and decreased immune responsiveness. Designing the optimal host milieu for tumor regression not only will need to replace deficient antitumor cytokines such as IL-2, IL-12, and interferon γ but will also need to decrease expression of these proinflammatory cytokines. A potential target that regulates cytokine expression is the transcription factor nuclear factor kappa B (NF-κB). NF-κB acts on the promoter site for many of these proinflammatory and proangiogenic cytokines to increase their expression. As a result, inhibiting NF-κB may be a promising immunotherapeutic modality for patients with HNSCC. A proteosome inhibitor of NF-κB (PS-341) was shown in preclinical studies to decrease production of these cytokines. This is expected to result in decreases in tumor growth, cell survival, and angiogenesis of HNSCC. Although, phase I clinical trials in patients with HNSCC have just begun, proteosome inhibitor–based therapy has shown potent antitumor results in phase I/II studies in other tumors.

Modifying Antigen Presentation

Optimizing the presentation of tumor antigens to the immune system can reduce their tumorigenicity and promote immune recognition and destruction of tumors. This objective can be accomplished in two ways: increasing the expression of MHC molecules and of costimulatory molecules on the surface of tumor cells, and improving antigen presentation by dendritic cells. Because T-cell activation requires two signals, one from interaction between MHC with the T-cell receptor and the other between costimulatory molecules with their counterreceptors on the T cell, providing these two key signals to tumor cells may augment T-cell proliferation and cytotoxicity.

Gene therapy has been used to transfect the alloantigen HLA-B7 (an MHC class I molecule) to tumors in patients with HNSCC who failed conventional therapy and whose tumors did not express this MHC antigen. Partial responses or stable disease was obtained in these patients after gene therapy. Biopsy results showed evidence that HLA-B7 expression had been achieved in these tumors. This study demonstrated the feasibility of enhancing tumor antigen presentation in patients with HNSCC. A phase II multicenter study is under way to investigate the efficacy of this treatment in a more adequate sample size.

An inhibitory signal for T-cell activation is generated when costimulatory molecules (CD80, CD86) on the APC interact with the CTLA-4 receptor on T cells. Unlike signals generated from CD80/CD28 interactions, signaling through CTLA-4 leads to decreased production of IL-2, cyclins, and cyclin-dependent kinases and restricts T-cell proliferation. The dynamic interaction between T-cell receptor, CD28, and CTLA-4 signals determines the outcome of T-cell activation. Phase I–II trials using antibodies against human CTLA-4 in prostate cancer and melanoma are ongoing, and research in patients with HNSCC will soon follow.

Another immunotherapeutic strategy that has received much attention is the use of dendritic cells to enhance tumor antigen presentation to the immune system. Recent evidence suggests that dendritic cells primed with tumor antigens can stimulate the regression of solid tumors such as renal cell carcinoma and melanoma. Generating sufficient numbers of dendritic cells to drive large clinical-scale expansions of lymphocyte cultures remains a significant challenge for dendritic cell–based therapies. Priming dendritic cells to generate tumor-specific T lymphocytes has also been a challenge in the development of dendritic cell–based therapies for HNSCC, because HNSCC–specific peptide antigens are largely unknown. Because it has been demonstrated that dendritic cells can efficiently acquire antigens from apoptotic tumor cells and cross-prime cytotoxic T cells, research on dendritic cell–based therapy in HNSCC is showing better outcomes in preclinical studies. Encouraging results from these studies have warranted two phase I clinical trials that are under way. In these trials, dendritic cells are derived from patients and are pulsed with irradiated tumor cells.

Transferring Immune Cells

Adoptive immunotherapy or *adoptive cell transfer* refers to the isolation of antigen-specific cells, their expansion and activation ex vivo, and their autologous administration to the host. Several cell types have been used for adoptive immunotherapy: T cells (peripheral or TIL), LAK cells, and dendritic cells. Preclinical studies on adoptive immunotherapy have provided the earliest evidence that established tumors can be induced to regress. However, several factors have delayed successful clinical studies using adoptive immunotherapy. Although it has been shown that CD8+ T cells are essential for antitumor effects in many models, the number and effectiveness of cells transferred during adoptive immunotherapy are directly correlated with treatment efficacy. The addition of T-cell growth factors such as IL-2 has significantly increased the effectiveness of transferred cells. Moreover, circumventing the immunosuppressive effects of regulatory T cells may improve immune responses to cancer cells. The feasibility of systemic adoptive T-cell immunotherapy in 15 patients with unresectable HNSCC was demonstrated in a phase I trial using autologous irradiated tumor cells and T cells. However, only two patients showed clinical responses. Although adoptive T-cell immunotherapy continues to be an active area of interest for other cancers, it is not thought to be an ideal target in HNSCC. This is largely because these cells have been found to be defective (decreased killing and decreased proliferation) in HNSCC. Clearly, this obstacle must be overcome to generate effective responses to adoptive transfer of T cells.

LAK cells are NK cells or T lymphocytes that have not been exposed previously to antigen and subsequently are activated using IL-2. Use of these cells in murine models has shown significant tumor regression. Autologous LAK cells have been administered to patients with HNSCC in multiple phase I studies with or without IL-2 or GM-CSF. The infusions were well tolerated; however, partial or minimal responses were observed at the primary tumor sites. The limited clinical benefit evidenced in these clinical trials emphasizes the need for continued research into strategies to improve tumor antigen–specific therapies against HNSCC.

In summary, no specific immunotherapeutic intervention has shown great promise in clinical studies of patients with HNSCC. Although tumor antigens on HNSCC have been recognized and tumor antigen–specific immune responses have been demonstrated, these immune responses have been ineffective in rejecting newly appearing and established tumor cells. The field of tumor immunology of HNSCC has advanced significantly in the past 30 years, although many obstacles remain that impede successful immune treatment of these cancers. A better understanding of the basis of immunosurveillance and tumor escape, antigen presentation to cytotoxic cells, mechanisms of host immunosuppression, and the complex cytokine network between tumor and host is needed. Deciphering the ideal strategy for immune treatment in these patients remains a significant challenge as well and may ultimately require a cohesive approach involving surgical debulking to reduce the immunosuppressive cytokine release by tumor and systemic or local immune therapy to augment antigen presentation and provide cytokines and other necessary signals for effective immune killing. As the effectiveness of promising new immune agents and strategies is established in clinical trials in patients with HNSCC, antigen-specific immunotherapy may be considered as an alternative modality for adjuvant treatment of patients with this disease.

Suggested Readings

Medawar PB. The use of antigenic tissue extracts to weaken the immunological reaction against skin homografts in mice. Transplantation. 1963 Jan;1:21–38

Ridge JP, Fuchs EJ, Matzinger P. Neonatal tolerance revisited: turning on newborn T cells with dendritic cells. Science. 1996 Mar 22;271(5256):1723–1736

Immunobiology of Head and Neck Squamous Carcinoma

Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol 2002;3(11):991–998

Hoffmann TK, Muller-Berghaus J, Ferris RL, Johnson JT, Storkus WJ, Whiteside TL. Alterations in the frequency of dendritic cell subsets in the peripheral circulation of patients with squamous cell carcinomas of the head and neck. Cancer Res 2002; 8(6):1787–1793

Reichert TE, Strauss L, Wagner EM, Gooding W, Whiteside TL. Signaling abnormalities, apoptosis, and reduced proliferation of circulating and tumor-infiltrating lymphocytes in patients with oral carcinoma. Clin Cancer Res 2002; 8(10):3137–3145

Whiteside TL. Immune cells in the tumor microenvironment: mechanisms responsible for functional and signaling defects. Adv Exp Med Biol 1998;451:167–171

Wustrow T, Issing W. Immune defects in patients with head and neck cancer. Anticancer Res 1993;13:2507–2519

Immunotherapy of Head and Neck Squamous Carcinoma

Gleich LL, Gluckman JL, Armstrong S, Biddinger PW et al. Alloantigen gene therapy for squamous cell carcinoma of the head and neck: results of a phase-1 trial. Arch Otolaryngol Head Neck Surg. 1998 Oct;124(10):1097–1104.

Hoffmann TK, Bier H, Whiteside TL. Targeting the immune system: novel therapeutic approaches in squamous cell carcinoma of the head and neck. Cancer Immunol Immunother 2004 Dec;53(12):1055–1067

Resser JR, Carbone DP. Immunotherapy of head and neck cancer. Curr Opin Oncol 1998;10(3):226–232

Van Waes C, Chen Z, Callister M, et al. Cytokines in the immune pathogenesis and therapy of head and neck cancer. In: Veldman JE, Passali D, Lim JD, eds. New Frontiers in Immunobiology. Kugler Publications, The Netherlands; 2000:233–243

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. The following are tumor-induced mechanisms to subvert host anticancer immune responses:
   A. Alterations of major histocompatibility complex class I and tumor antigen expression
   B. Increased surface expression of costimulatory molecules
   C. Secretion of soluble immunosuppressive factors
   D. All of the above
   E. A and C only
2. Which of the following is involved in cell-mediated immunity in HNSCC?
   A. Cytokines
   B. Antibodies
   C. NK cells
   D. Macrophages
   E. A, C, and D only
3. Which immunologic abnormality in cell-mediated immunity is not seen in patients with HNSCC?
   A. Dysfunction of CD4+ T cells
   B. Increased NK cell activity
   C. Decreased proliferation and cytotoxicity of CD8+ T cells
   D. Dendritic cell dysfunction
   E. All of the above
4. Which of the following cytokines is being investigated as an immunotherapeutic agent in HNSCC?
   A. IL-2
   B. IL-12
   C. IL-1 α
   D. A and B
   E. All of the above
5. Which cells have been used in immunotherapy because of their ability to present antigens to the immune system?
   A. CD8+ T cells
   B. CD4+ T cells
   C. Dendritic cells
   D. NK cells
   E. All of the above

Chapter 11

CLINICAL RADIATION BIOLOGY AND RADIOTHERAPY

STEVEN R. ISAACSON AND LANNY GARTH CLOSE

CLINICAL APPLICATIONS OF RADIATION THERAPY

Radiation therapy may be integrated into the management of head and neck cancer in a variety of ways. The manner in which it is applied depends on many factors, including the stage of disease, extent of primary and nodal involvement, presence of clinical and pathological prognostic factors, opportunity for organ preservation, desire to preserve form and function, overall patient condition, and the skill and experience of the specialists involved.

The combination of surgery plus radiation should be considered under certain circumstances, which include the following:

1. Close or positive surgical margins at the primary site
2. Large ($\geq$ T3) primary lesions or lesions at sites where complete resection is problematic (skull base, infratemporal fossa, brachial plexus, internal carotid artery)
3. Cervical node metastasis beyond primary lymphatic drainage area or multiple positive nodes, multiple levels of positive nodes, or extracapsular nodal disease

PREOPERATIVE RADIATION

The planned use of preoperative radiation therapy takes advantage of several theoretical circumstances. The blood supply to a tumor preoperatively may be significantly better than after surgery. This makes the clonogens more responsive to radiation. Lesions felt to be unresectable or at high risk for positive or uncertain margins may be converted to a more resectable status. Also, the possibility of tumor embolization during surgery resulting in distant metastasis may be decreased.

Disadvantages of preoperative radiation include the fact that the dose of radiation must be limited (~50 Gy) to avoid complications. Also, there is the decreased benefit of surgical staging. Finally, the patient may have an unfounded belief that no further treatment (surgery) is necessary if there is a complete response. Preoperative radiation should be considered if the neck has been violated (previous incisional biopsy), if a delay in surgical recovery is anticipated, or, in the case of large cervical metastases, if the primary site can be treated with irradiation.

Surgery following planned preoperative radiation is relatively safe, in sharp contrast to surgical salvage following failed, full-course irradiation with intent to cure. In the case of planned preoperative radiation, the dose of radiation is relatively low (50 Gy over 5 weeks to mucosal surfaces, 55 Gy over 5½ weeks to lymphatics), and the interval between irradiation and surgical intervention is relatively short (1 week per 10 Gy of radiation). Surgery is carefully timed to follow the period of acute inflammation and precede the delayed effects of irradiation (increased fibrosis, decreased vascularity, and decreased cellularity). Surgical salvage follows high-dose irradiation (~70 Gy over 7 weeks) and invariably is performed 4 months or more after the last dose of irradiation, resulting in poor healing and an increased incidence of complications due to the delayed effects of irradiation.

A unique problem arises when the patient received definitive irradiation to the primary combined with planned preoperative radiation to cervical metastases. To avoid complications due to the delayed effects of irradiation, a neck dissection should be performed within 6 weeks following radiation therapy. If a biopsy of the primary site is performed at this time, cancer cells are likely to be found in the specimen. Unfortunately, the viability of these cancer cells cannot be determined at this early interval. Thus the surgeon is faced with a difficult decision regarding the primary site: should the patient undergo extirpative surgery at the time of neck dissection or should the surgeon follow the patient for clinical evidence of a residual (recurrent) primary lesion?

POSTOPERATIVE RADIATION

Postoperative radiation is given when the attempt at surgical extirpation has been completed and healing has begun. Its purpose is to address clinical and pathological findings that are known to lead to failure at the primary and regional node sites. Postoperatively, radiation therapy can be given at greater doses of radiation to less tumor than may be given preoperatively. Several proposed indications for postoperative radiation include positive (close) margins, perineural spread, lymphovascular invasion, bone invasion, extension into soft tissues, multiple involved lymph nodes, positive lymph nodes greater than 3 cm, and extracapsular spread in lymph nodes of any size.

In the United States today, the standard of care for the use of combination surgery and irradiation is surgery followed by postoperative irradiation. Despite the current popularity of this treatment, one should consider planned preoperative irradiation as a safe, effective treatment when given at the correct dose (50 Gy to mucosa, 55 Gy to lymphatics) and followed by surgery at the correct interval (1 week delay per 10 Gy). In a large multiinstitutional double-blinded, randomized prospective protocol treating squamous cell carcinoma of the head and neck, preoperative external beam radiation therapy compared favorably to postoperative external beam radiation therapy in terms of complications (no difference) and patient 5-year survival (no statistical difference). Patterns of cancer failure were noted (decreased local/regional failure after postoperative external beam radiation therapy, decreased distant metastases after preoperative external beam radiation therapy).

If postoperative irradiation is given, the interval between surgery and the initiation of irradiation is critical. Despite some evidence to the contrary, most retrospective studies (no prospective study has been performed) indicate that a delay no greater than 4 to 6 weeks following surgery is preferable. If a neck dissection is performed prior to irradiation with intent to cure the primary combined with postoperative irradiation to the neck, the delay following surgery until the initiation of radiation therapy should not be greater than 2 weeks.

RADIOBIOLOGY OF RADIATION THERAPY

The delivery of ionizing radiation in the treatment of tumors also affects normal tissues. The biological consequences of radiation upon these tissues are generally understood. Rational attempts at enhancing the biological effects of the radiation therapy are being formulated with the understanding of these consequences. Among these are not only physical methods of biological enhancement but also therapeutic methods that involve more biologically correct ways of delivering the daily and total doses of radiation therapy.

Radiation striking the tissue of a patient affects the biology of both normal and tumor tissues. Ionizing radiation causes both direct and indirect effects on biological targets. The deoxyribonucleic acid (DNA) of a cell may be affected directly by the secondary electrons generated as ionizing radiation interacts with tissue. The radiation also may have an indirect effect due to the formation of free radicals; these free radicals in turn cause most of the chemical damage to the theoretical target, DNA. This is the predominant reaction that we are concerned with in radiation therapy of the head and neck. In addition, there are several other cellular functions that are disrupted by radiation-induced damage. This damage may be modified by oxygen concentration, temperature, and other intracellular components. The two most important biological consequences of radiation therapy are the loss of reproductive ability and cellular function. In the treatment of head and neck cancers, it is the loss of clonogenic reproductive ability that is the most important consequence of ionizing radiation. This term describes the loss of the cell's ability to continue producing normal daughter cells or, in the case of tumors, additional clonogens.

The relationship between the dose of radiation and the loss of reproductive ability is demonstrated by cell survival curves. These curves demonstrate the ability to survive a given dose of radiation. Both normal and tumor tissues have characteristic shapes for cell-survival curves. The most frequently described relationship for mammalian tumor cells is described by the equation

$$S = 1 - (1 - e^{-D/Do})^{n}.$$

In this equation, which demonstrates surviving fraction (S), as a function of dose (D) and predicts a single-hit multi-target model, there is demonstrated an initial part of the curve that is a "shoulder." The extrapolation number "n" is a measure of the "shoulder" of the survival curve. In this "shoulder" is the amount of damage that the cell can incur and must accumulate before cellular lethality. The term Do is the final slope of the curve and represents the dose required to reduce survival from 0.1 to 0.037 or from 0.01 to 0.037.

Another model that has become significant in its popularity is the linear-quadratic model. This assumes that cell killing has two components; one is proportional to the dose (D), and the other is proportional to the dose squared (D^2). This equation is:

$$\text{Log}_e S = \alpha D + \beta D^2.$$

Survival is represented by S. The linear single-hit killing component is represented by α. The quadratic multiple-hit component is represented by β. Tumors that are clinically responsive to radiation therapy demonstrate a high α:β ratio. Tumors with low α:β ratio are frequently radioresistant. Early-reacting tissues in this equation also give rise to the concept of acutely reacting tissues (e.g., tumor cells), and mucosal tissues and late-responding tissues such as nerve, muscle, bone, and fat. The linear-quadratic model forms the basis for the development of newer fractionation strategies that allow the α:β ratios of tumors and normal tissues to be exploited to the advantage of controlling tumors while keeping complications (late-responding tissue effects) to a minimum.

There are several factors that will influence the radiation survival curves depicted by the previous formulas. These include the oxygen content of the cells, the repair of the cellular processes following either sublethal or potentially lethal injury (e.g., the position in the cell cycle), the potential for clonogen proliferation, and the inherent radiosensitivity of the cell. All of these factors have consequences in the way the tissues respond to radiation. The tissues are broadly divided into early- or acute-responding tissues (of which mucous membrane and tumor are included) and late-responding tissues, in which the effects of radiation will occur many weeks or months after the radiation has been given. Late-responding tissues are predominantly those in which we see complications of radiation therapy and include muscles, nerve, bone, and fat. Available radiobiological evidence leads us to believe that these late-responding tissues act significantly differently in response to the amount of radiation given in each fraction of radiation. It is the size of this individual fraction (e.g., 1.8–2.0 Gy qid) that influences these late tissue effects the most.

The response of tumors to radiation is highly dependent on the size of the tumor as well as the several concepts that play a role in the eradication of tumors. The processes of repair of radiation-induced damage, repopulation of clonogens, reoxygenation of hypoxic tumors, and redistribution of cells into radiation-sensitive positions in the cell cycle are very important in the control of tumors. These constitute the four R's of radiobiology. These four R's are the basis for the clinical delivery of radiation therapy.

FRACTIONATION OF TREATMENT IN RADIATION THERAPY

The aforementioned biological principles of radiation therapy are most important in developing a rationale for the fractionation of treatment in radiotherapy for head and neck malignancy. *Fractionation* is a term used to describe the clinical manner in which the daily dose of radiation is given. It remains one of the foundations of contemporary radiation therapy for head and neck carcinoma.

There is, however, no universally accepted standard for fractionation. The existence of many "conventional" schemes has been influenced by personal opinion, socioeconomic factors, equipment availability, and historical and traditional prejudices. In the United States, the standard is 1.8 to 2.0 gray (Gy) or 180–200 centigray (cGy) once daily, Monday through Friday, over approximately 7 weeks. In Canada, treatment is 250 cGy per day 4 days a week over 5 weeks. In England, the therapy may extend through the weekend at higher fraction totals per day, for example, 150 cGy 3 times a day.

Recently, an attempt to use radiobiological therapy as it pertains to the differences in early- and late-reacting tissues has come into prominence. Examples of early-reacting tissues are tumor cells, mucous membrane cells, and gastrointestinal mucosa. Late-effects tissues include, but are not limited to, those in which we see complications of therapy such as muscle, nerve, and bone.

TABLE 11–1 FRACTIONATION

Biological basis
- α:β ratio
- Tumor cell proliferation

Objective
- Exploit inherent tissue differences
- Treat in a shorter time period

Currently, altered fractionation (nonstandard) reforms are being designed to account for this clonogen growth or to exploit the different biological responses of acute (tumor) and late- (complication) reacting tissues. These two biological strategies are the main foundations of contemporary fractionation policies (**Table 11–1**).

RATIONALE FOR FRACTIONATION

It has been demonstrated over the last several years that tumor cell growth actually accelerates exponentially during the course of radiation therapy. It is therefore not unreasonable to deliver radiation therapy in a more accelerated manner so as to overcome this exponential growth. These two biological strategies, differentiating early and late effects and overcoming tumor proliferation, are the main foundations of contemporary fractionation policies.

Several physical factors are important in the discussion of any fractionation regime. Dose per fraction, total dose (number of fractions), overall time, and time interval between the delivery of radiation fractions are the most critical of these factors. The dose per fraction is most commonly described as being either a rad (old) or a gray (new). A rad is a measure of energy absorption of 100 erg/g. A gray is energy absorption of 1 Joule/kg. One gray is equal to 100 cGy (rads). One cGy equals 1 rad. Decreases in the size of the dose per fraction spares late-responding normal tissue (e.g., nerves, bone, and muscle) more than it spares tumor cell killing. This is one of the implications of the linear-quadratic model (α:β ratio).

Overall time is also important for any given effect; total dose delivered to the tumor has to be increased if overall time in which the treatment is given is increased. Small fractions are less effective than large fractions and require compensation for proliferation in normal tissue and, of course, tumors. In the treatment of head and neck cancer it is often a possibility that the overall time in which the treatment is given has to be prolonged. In this case, the total dose must be increased. Small fractions are less effective than larger fractions in doing this, and one must compensate for proliferation both in normal tissues and in tumor tissues. It is important, however, not to consider time alone. Normal tissue effects in early-reacting tissues start compensatory proliferation in 2 to 4 weeks after the start of radiation therapy. Late-reacting tissues, which are the origin of complications, have no proliferation during the weeks of radiation therapy. It is important to try to keep the overall treatment period as short as reasonably possible. However, we try not to do this by greatly increasing the size of the daily fractions. Late effects (complications) are fraction size dependent. This is another implication of the linear-quadratic model. Prolonging the overall time decreases damage to acute tissues and therefore makes the treatment more tolerable to patients with less acute side effects. However, this dose does not cause a decrease in the damage to late tissues. Prolonging the overall treatment time will result in less damage to the tumor cells that we are trying to control and therefore will result in less local control. Shortening the overall treatment time will increase the damage to acute tissues with no increase in damage to the late tissues unless the dose per fraction is increased significantly in its daily size. Shortening also results in an increased tumor kill.

In conventional treatment the total dose is determined by type of tumor and volume being treated as well as the tolerance of normal tissue in the volume. As stated earlier, altered fractionation schemes either try to exploit the difference between early- and late-reacting tissues or take into consideration the new tumor cell kinetic information that has shown us tumors will become rapidly exponential in their growth while under treatment. Hyperfractionation, accelerated fractionation, and dose escalation (e.g., concomitant boost) are several strategies that have become popular throughout the world (**Table 11–2**).

Hyperfractionation is administered by a decrease in the size of the administered fraction. There is an increase in the total number of fractions and an increase in the total dose. However, the overall treatment time is maintained at what would have been expected from a conventional treatment. The most classic example of this is the delivery of 1.2 gray fractions twice a day for approximately 7 weeks. This is the same 7-week period in which conventional fractionation at 2 Gy once a day may have been delivered. Using the linear-quadratic model as the basis for computing the biological effects of doing this results in an 84 Gy tumor dose in the same amount of time that a 70 Gy tumor dose would have been delivered conventionally. Theoretically, this smaller fraction size allows sparing of late tissue damage while an increased total dose can be given to the tumor.

Accelerated treatment seeks to decrease the overall time it takes to deliver a course of radiation therapy. During this time the fraction size, fraction number, and

TABLE 11–2 FRACTIONATION IN RADIATION THERAPY

Scheme	Total Dose	Overall Time	Fraction Size	Fraction Number
Conventional	60–70 Gy	6–7 weeks (Std)	2 Gy (Std)*	30–35 (Std)
Hyperfractionated	Std/†	–	–	–
Accelerated	Std	‡	Std (SI)	Std (SI)
Concomitant boost**	Std (SI)	Std	Std (SI)	

*1.8–2.0 Gy
**A type of accelerated treatment
†When total dose is not increased, called quasi-hyperfractionated.
‡When treatment break is given, called quasi-accelerated.
SI, slightly; Std, standard

total dose are either maintained or very slightly altered. For example, if one were to deliver a treatment once a day for 5 weeks, from Monday through Friday, this treatment might deliver the same 25 individual treatments but with no break for Saturday or Sunday. Thus the 5 weeks required for treatment would become 3½ weeks. This shortening of the overall treatment time theoretically defeats the tumor proliferation that takes place during the course of the standard treatment of 5 to 7 weeks.

A combination of both of the previously mentioned strategies is an accelerated hyperfractionation strategy. This combines the shortening of the overall treatment time to defeat tumor proliferation and the smaller fraction size to spare late tissue damage. Examples of this include the delivery of 35 fractions at 1.5 Gy delivered 3 times a day in 12 days or 1.6 Gy twice a day at 24 fractions in 3.4 weeks with a break to allow for the early-reacting tissue to decrease its bothersome response, followed by resumption of treatment at 1.6 Gy twice a day to the 64 to 67.2 Gy level.

Another strategy is a type of dose escalation or concomitant boost. With this strategy, a small booster field is delivered within a larger field on the same day. These treatments are delivered at least 6 hours apart, and, although approximately the same total dose as a standard treatment is given, there is a change in the fraction sizes delivered once in the morning and once in the afternoon. The diminished size of the booster dose decreases the volume to which radiation is given and thus decreases the effects early-reacting tissues such as mucosa while having a great effect on the tumor tissue. Treating the tumor tissue twice in a single day at near standard fractionation overcomes the tumor cell proliferation.

Another advantage of accelerated fractionation schedules is a reduction of overall duration of treatment. This allows tumor proliferation to be overcome without causing additional injury to late-reacting tissues. There are several caveats to be considered in the delivery of altered fractionation schemes based on the radiobiological rationales of late/early tissue sparing and overcoming tumor cell proliferation. Although several pilot studies using hyperfractionation and accelerated fractionation suggest a therapeutic advantage, there is no convincing evidence of an improvement in the therapeutic ratio. Several trials are ongoing.

Both hyperfractionation and accelerated fractionation are associated with increased acute normal tissue reactions that may be dose limiting. This is certainly the case when fractionation is delivered three times a day.

Altered fractionation regimes may also be associated with unexpected late normal tissue complications. These may be due to the normally expected sequelae of treatment or may actually be related to the size of the dose per fraction, the number of fractions delivered in a single day, and the time between these fractions.

Most normal tissue reactions can be held to a minimum if an interfraction interval of 6 hours is allowed. The only exception to this may be the spinal cord, where even a 6-hour interfraction interval may not be enough to overcome the effects.

SELECTED READINGS

Amdur RJ, Parsons JT, Mendenhall WM, et al. Postoperative rradiation for squamous cell carcinoma of the head and neck: an analysis of treatment results and complications. Int J Radiat Oncol Biol Physiol 1989;16:25–36

Byers RM, Clayman GL, Guillamondequi OM, et al. Resection of advanced cervical metastasis prior to definitive radiotherapy for primary squamous carcinomas of the upper aerodigestive tract. Head Neck 1992;14:133–138

Fein DA, Lee WR, Amos WR, et al. Oropharyngeal carcinoma treated with radiotherapy: a 30 year experience. Int J Radiat Biol Physiol 1996;34:289–296

Hall E. Radiobiology for the Radiologist. 4th ed. Philadelphia: JB Lippincott; 1993

Kahn F. The Physics of Radiation Therapy. Baltimore: Williams & Wilkins; 1993

Marcial VA, Gelber R. Kramer S, et al. Does preoperative irradiation increase the rate of surgical complications in carcinoma of the head and neck? A Radiation Therapy Oncology Group Report. Cancer 1982;49:1297–1301

Mendenhall WM, Parsons JT, Stringer SP, et al. Squamous cell carcinoma of the head and neck treated with irradiation: management of the neck. Sem Radiat Oncol 1992;2:163–170

Peters LJ, Ang KK. The role of altered fractionation head and neck cancers. Sem Radiat Oncol 1992;2:180–194

Taylor JMF, Withers HR, Mendenhall WM. Dose time consideration of head and neck squamous cell carcinomas treated with irradiation. Radiother Oncol 1990;17:95–102

Tupchong L, Phil D, Scott CB, et al. Randomized study of preoperative versus postoperative radiation therapy in advanced head and neck carcinoma: long-term follow-up of RTOG Study 73–03. Int J Radiat Oncol Biol Physiol 1991;20:21–28

Withers HR, Taylor JMF, Maciejewski B. The hazard of accelerated tumor clonogen repopulation during radiotherapy. Acta Oncol 1988;27:131–146

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. When hyperfractionation is used to treat head and neck cancer with curative intent, which of the following statements is true?
   - A. The daily fraction size is "standard" in size.
   - B. The overall time is shortened.
   - C. The overall dose is increased.
   - D. Acute toxicity is less.
2. When accelerated treatment schemes are used to treat head and neck cancer with curative intent, which of the following statements is true?
   - A. The daily fraction size is increased so the treatment will finish faster.
   - B. The total dose is increased.
   - C. The toxicity is lessened.
   - D. The overall treatment time is significantly decreased.
3. Hyperfractionation is used to
   - A. Rapidly finish treatment
   - B. Exploit the radiobiological difference between acutely and late-reacting tissues
   - C. Achieve a higher total dose
   - D. Lessen toxicity
4. The point on a survival curve where the single-hit component equals the multiple-hit component is known as
   - A. Nirvana
   - B. Therapeutic index
   - C. Survival equilibrium
   - D. α:β ratio
5. The biological rationale behind accelerated fractionation schemes shortens the overall treatment time dramatically because
   - A. Tumor doubling time is slow, and therefore damage can accumulate.
   - B. Large-dose per fraction treatment is therapeutically better.
   - C. Patients become fatigued following long periods of radiation.
   - D. Tumor clonogens start to grow rapidly while under treatment.

Chapter 12
ENVIRONMENTAL EFFECTS ON THE UPPER AIRWAY

Andrew Blitzer

Many inhaled materials from the environment have a detrimental effect on the nose, paranasal sinuses, larynx, and trachea. Oxidant gases, in particular, produce a variety of changes in the airway, increasing in severity from injury and exfoliation of ciliated cells to severe airway damage, resulting in exposure of the basement membrane of both mature and undifferentiated cells. The airway responds to these insults with adaptive mechanisms, including decreased sensitivity of airway receptors, decreased airway permeability, increased number of secretory cells, and increased mucus production. Additionally, damage to the ciliated cells results in an increased need for coughing to clear both inhaled particles and excess secretions.

TOXICITY

The toxicity of each inhalational agent depends on its intrinsic toxicity, its absorption and transport to the target organ, its ability to biotransform into a more or less toxic substance, and its ability to bind to essential macromolecules. The pathophysiological effects will be mediated by multiple factors, including the concentration of the agent and the exposure time, the gas solubility, and certain environmental characteristics, such as temperature, humidity, and airflow. The simultaneous presence of other gases or pathogens can also mediate the severity of the injury. At the cellular level, the acute effect of toxic chemicals on the respiratory epithelium results in increases in intracellular calcium, dysfunction of sodium regulation, vacuolization of the cytoplasm, dilation of the endoplasmic reticulum, condensation of matrix proteins, and apical membrane blebbing.

CHEMICAL POLLUTANTS

Common examples of damaging gases in the indoor environment are carbon monoxide and nitrogen dioxide. These ubiquitous pollutants are by-products of cigarette smoking and the use of certain heating systems. Excessive exposure to nitrogen dioxide may cause a reduction in cellular defense mechanisms and may lead to dyspnea, restlessness, coughing, wheezing, nausea, and vomiting.

Numerous other chemicals have been shown to damage the respiratory system. Formaldehyde, which can be released by certain insulation materials or be present in high concentrations in anatomy and pathology laboratories, can cause eye, nose, throat, and lung irritation. Chronic exposure has been associated with sinonasal, nasopharyngeal, and hypopharyngeal cancers. Additionally, an increased rate of leukemias is found in embalmers, anatomists, and pathology technicians. Benzene, still used as a solvent in many laboratories, can cause bone marrow suppression. Volatile organic compounds associated with chronic airway irritation may come from common office equipment such as copiers and computer printers. Exposure to the above volatile compounds has also been associated with aplastic anemias and leukemias. Other polychlorinated solvents, notably aromatics (e.g., xylene), and polychlorinated solvents (e.g., carbon tetrachloride, tetrachloroethylene, and trichloroethylene), have all been associated with liver degeneration, hepatitis, and renal disease.

Excessive exposure to isopropyl alcohol has been linked to cancer of the paranasal sinuses and larynx. Bis(chloromethyl)ether exposure has been found to produce esthesioneuroblastomas in a rat model. The organic solvents may also cause chronic encephalopathies and peripheral neuropathies. Toluene and xylene are both associated with menstrual disorders. Ethylene oxide has been found to cause chronic airway irritation. Inhalation of hydrazine has been associated with polyps, carcinoma, and, possibly, amyloidosis.

Chronic inhalation of tobacco smoke has been proven to cause significant airway damage, ranging from sinusitis and laryngitis to carcinoma of the entire respiratory tract. Inhalation of marijuana in heavy smokers has been shown to cause rhinitis, pharyngitis, and laryngitis. Other contaminants in smoke (such as *Aspergillus*) may compound the problem. Inhalation of cigarette smoke may produce atypism and cellular hyperplasia, loss of cilia, and possible carcinogenesis. Cocaine inhalation causes chronic rhinitis and pharyngitis, and its effects are believed to be related both to its vasoconstrictive properties and to the presence of adulterants.

Chloroform is now rarely used; its well-described dangers have prompted the U.S. Food and Drug Administration to ban its use in drugs and cosmetics. However, it is still used commonly in certain endodontic procedures. Chronic exposure to chloroform, especially in small, confined areas, places these dental workers at increased risk for lung and liver disease.

Anesthetic gases represent a major environmental hazard in the operating room. Several studies performed in the 1970s concluded that the rate of spontaneous abortions was substantially higher in female anesthesiologists compared with female physicians working outside the operating room. Furthermore, the incidence of congenital anomalies seen in children of both male and female anesthesiologists was higher than in the control group of physicians. The presence of malignancy was similar in both anesthesiologists and controls. Liver disease, however, was more prevalent in male anesthesiologists.

It has been shown that concentrations of nitrous oxide as low as 50 ppm, either alone or in combination with 1 ppm of halothane, result in decreased behavioral performance. Although the data have been disputed by other researchers, they show that anesthetic gases may have an insidious effect on the operating room personnel. Gas-scavenging techniques have been employed to minimize exposure. New data, however, suggest that even with the use of scavenging techniques, chronic inhalation of trace concentrations of anesthetic gases may be harmful.

Another potentially harmful exposure in the operating room is related to the increased use of laser surgery. With vaporization of lesions, a smoke plume (the so-called laser plume) is produced, composed mostly of water vapor and carbon particles. However, in lesions such as papillomas, live viral particles and bacteria have been found. Measuring on average 0.31 μm (range 0.1–0.8 μm), these pose a risk of inhalation to operating room personnel, given the fact that most surgical masks do not trap particles this size. Thus great efforts need to be taken to scavenge the laser plume using specialized ventilation/evacuation systems.

Acrylic monomers, commonly used for cranioplasty, orthopedic, or dental procedures, may cause airway irritation and hepatotoxicity. As inhalants, acrylic monomers represent a particular risk to dental personnel because they are aerosolized during dental procedures. Chronic exposure can cause central nervous system (CNS) disturbances, including tremor, ataxia, nervousness, irritability, paresthesias, and visual changes.

Other inhalational agents noxious to the airway are metals, such as mercury, magnesium, zinc, cadmium, and nickel. Mercury is particularly risky to dental personnel. Its aerosolized form, used in dental procedures, can lead to chronic CNS effects similar to those already described for acrylates. Magnesium is also irritating to the airway and may cause cough, rhinitis, and increased mucus production. Zinc is used to galvanize iron and is found in paints, pigments, glazes, enamels, and certain paper manufacturing processes. Acute inhalation of zinc fumes may cause "metal fume fever" and significant airway irritation. Cadmium is also terribly noxious to the airway. The major nonindustrial source is cigarette smoke. Exposure is associated with cancers of the respiratory tract, lung, and prostate. Cadmium fume inhalation resulting from exposure to industrial processes such as stainless steel production, plating, tanning, and pigment and wood preservative manufacturing may cause nasal irritation and septal perforations, as well as chronic rhinosinusitis, pharyngitis, laryngitis, and lung cancer. Nickel fumes are associated with chronic airway irritation and lung cancer.

ENVIRONMENTAL INHALANTS

Environmental inhalants can cause a constellation of constitutional symptoms, ranging from mild to severe. Most large buildings use water-cooled ventilation systems, which are often contaminated with molds, mildews, and bacteria. Central ventilation ensures that these contaminants are efficiently pumped throughout the entire building. The best known example is that of the *Legionella pneumophila* epidemic of 1976, which caused 189 cases of pneumonia and 29 deaths in residents of a Philadelphia hotel. Legionella was found to have grown in the stagnant water of a cooling tower and to have been spread as an aerosol within the ventilation system.

Humidifiers associated with cooling systems are a source of air contaminants such as fungi and dust mites, as are air-conditioning filters. A common fungal organism spread in hospital ventilation systems is *Aspergillus*. The spread of fungal spores may be particularly dangerous for immunocompromised patients. However, even in immunocompetent hosts, fungal contamination of the ventilation systems has been associated with allergic alveolitis, rhinitis, and sinusitis. Tuberculosis has also been found in ventilation systems, and this could pose a particularly difficult problem with the emergence of new, multidrug-resistant strains. In addition to microorganism contamination, larger particles, such as animal hair and dander, are spread through building ventilation systems.

Besides the above contaminants, building inhabitants are exposed to the biocides placed in the ventilation systems to control the spread of microorganisms. Many of these agents have been found to be extremely irritating to the upper respiratory system.

SICK BUILDING SYNDROME

A constellation of physical, chemical, and biological environmental factors found in the indoor working environment has been identified and is commonly referred to as the "sick building syndrome." It has been reported to occur in 20 to 30% of all office workers, and it reportedly reduces their work productivity by up to 30%. Characteristic symptoms are often nonspecific: nasal irritation, rhinorrhea, nasal obstruction, chest tightness, dry and irritated eyes, dry and irritated throat, dry skin, rash, headaches, lethargy, and poor concentration. Several physical factors intensify the symptoms, including ambient temperature of more than 22°C, poor ventilation, and humidity greater than 70%. Artificial light also appears to play a role in this symptom complex. Environmental noise, negative ions, and inorganic dust have been associated with the syndrome. Chemical factors associated with sick building syndrome include exposure to secondhand smoking, formaldehyde vapors, volatile organic compounds (including office products such as photocopy ink), biocides, and other gases (e.g., carbon monoxide, carbon dioxide, nitrous dioxide, ozone, sulfur dioxide). Chronic exposures to biological contaminants such as mites, molds, bacteria, and fungi all appear to pay significant roles in the symptom complex. The sick building syndrome takes on added importance in hospital settings, because lethargy and poor concentration on the part of hospital personnel can harm patients.

To conclude, there are numerous indoor environmental exposures that can damage components of the respiratory system. Of note, hospitals present a special set of exposure risks, related to the use of inhalational anesthetics, lasers, and certain laboratory chemicals. In addition, the use of central cooling and ventilation systems exposes inhabitants to potentially toxic microorganisms and biocides. More work is needed to protect workers from the effects of indoor inhalational agents. As health care leaders, physicians need to assume active roles as patient and employee advocates and educators.

SUGGESTED READINGS

Amdur RJ, Parsons JT, Mendenhall WM, et al. Postoperative irradiation for squamous cell carcinoma of the head and neck: an analysis of treatment results and complications. Int Jradiat Oncol Biol Physiol 1989;16:25–36

Byers RM, Clayman GL, Guillamondequi OM, et al. Resection of advanced cervical metastasis prior to definitive radiotherapy for primary squamous carcinomas of the upper aerodigestive tract. Head Neck 1992;14:133–138

Fein, DA, Lee WR, Amos WR, et al. Oropharyngeal carcinoma treated with radiotherapy: a 30 year experience. Int J Radiat Biol Physiol 1996;34:289–296

Hall E. Radiobiology for the Radiologist. 4th ed. Philadelphia: JB Lippincott; 1993

Kahn F. The Physics of Radiation Therapy. Baltimore: Williams & Wilkins; 1993

Marcial VA, Gelber R, Kramer S, et al. Does preoperative irradiation increase the rate of surgical complications in carcinoma of the head and neck? A Radiation Therapy Oncology Group Report. Cancer 1982;49:1297–1301

Medenhall WM, Parson JT, Stringer SP, et al. Squamous cell carcinoma of the head and neck treated with irradiation: management of the neck. Semin Radiat Oncol 1992;2:163–170

Peters LJ, Ang KK. The role of altered fractionation in head and neck cancers. Sem Radiat Oncol 1992;2:180–194

Taylor JMF, Withers HR, Mendenhall WM. Dose time consideration of head and neck squamous cell carcinomas treated with irradiation. Radiother Oncol 1990;17:95–102

Tupchong L, Phil D, Scott CB, et al. Randomized study of preoperative versus postoperative radiation therapy in advanced head and neck carcinoma: long-term follow-up of RTOG Study 73–03. Int J Radiat Oncol Biol Physiol 1991;20:21–28

Withers HR, Taylor JMF, Maciejewski B. The hazard of accelerated tumor clonogen repopulation during radiotheraphy. Acta Oncol 1988;27:131–146

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 715.

1. "Sick building syndrome" includes
   A. Rhinorrhea and nasal obstruction
   B. Headaches
   C. Eye irritation
   D. Lethargy
   E. All of the above

2. Humidifiers have been associated with the spread of
   A. Aspergillosis
   B. Candidiasis
   C. Tuberculosis
   D. A and B
   E. B and C

3. Particularly noxious environments in the hospital include
   A. Operating rooms
   B. Pathology laboratories
   C. Photocopy centers
   D. All of the above
   E. None of the above

Chapter 13
HOW TO CONDUCT CLINICAL RESEARCH

STEVEN D. RAUCH

Scientific investigation is a process, the process of discovering the working of the natural world. Science answers the question, How? Science is a culture. It has its own language and etiquette, its own socialization and training routines, its own icons and articles of faith. Training to become a scientist is an indoctrination into this culture. Members learn certain ways of thinking, of posing questions, and of seeking their answers.

Clinical medicine is also a culture, albeit a different one from science. Imagine that you are a clinician, and a young scientist comes to see you. He has completed his doctorate and is nearing completion of his postdoctoral fellowship. He will be taking an academic position at a prestigious university within a year. He tells you that he has decided that his academic career will be more successful and fulfilling if he could have a clinical component in addition to his research activities. He would like to operate one afternoon a week. He also would like to spend 3 months with you in your practice to learn some surgery that he could do once he gets out of his fellowship. "Ludicrous," you say. You have spent 4 years in medical school and 5 years in residency to learn your craft. Despite the fact that the young scientist is intelligent and has excellent manual dexterity, there is no way he could master the art and science of clinical practice in 3 months, no way that he can become a member of the medical culture in that brief period of time.

Now imagine how it sounds to the senior scientist when a young resident, nearing completion of her training, comes requesting a few months of laboratory time so that she can "do science" one afternoon each week when she enters her academic job in a year or two. The scientist spent 7 years acquiring her PhD and 3 more years in her postdoctoral fellowship before she got her first grant support as an independent investigator. It is inconceivable that the resident, working part time in the laboratory for only a year or two, can learn how to pose a research question, develop a testable hypothesis, design an experiment, write a fundable research proposal, execute the experiment and analyze the results, and present or publish the findings. It is inconceivable that the resident could develop the kind of scientific intuition that only comes from years of experience, strong mentorship, and many frustrations and failures. It is also inconceivable that the resident could become a member of the science culture.

So, if the cultures of medicine and science are so different, why should members of one bother to try and learn about the other? The answer seems obvious. Each culture has something to offer that will enrich the other. Science can answer questions that will improve our ability to provide the best care to patients. Medicine can inform scientists of the clinical issues in greatest need of address and thereby enable scientists to focus on investigations of greatest relevance. The clinician-scientist becomes an ambassador between these two cultures, providing a conduit for transmission of ideas and information in both directions. A clinician-scientist becomes a valuable collaborator who can interact with "pure scientists" and "pure clinicians" to advance clinical care. Becoming a clinician-scientist, however, takes time and patience. There is no shortcut to membership in either culture. Whether the clinician spends several years full time or many years part time, development of research expertise will take as long as for the full-time scientist.

Clearly, not all physicians have the interest, desire, or stamina to become independent investigators. Is there a value in some less rigorous exposure to science? There are several good reasons for all physicians-in-training to have research experience. Most important is the need for clinicians to read and understand the medical literature. The only rational basis for clinical innovation is found in scientific contributions to the literature. If the clinician cannot read critically, he or she cannot decide whether to accept or reject new ideas from the research world. Critical reading depends on some familiarity with the process by which the new information was discovered. Awareness of clinical issues alone is inadequate. The critical reader must be able to assess validity of the experimental design, methodology of data collection and analysis, and interpretation of the results. Hands-on experience in conducting science is an invaluable asset in developing skills of critical reading. Another benefit of research experience during clinical training is to enable each trainee to make a lasting contribution to his or her field. Trainees may never enter another laboratory or conduct another experiment for the rest of their lives, but they will forever have the satisfaction of knowing there is an article in the literature and reprints on their shelves that represent their unique contribution to the specialty's body of knowledge. Finally, trainees deserve an opportunity to gain exposure to as many career options as possible. Because scientific investigation may be appealing to some, trainees will benefit from a chance to try it. Those with aptitude can be encouraged to seek further training in the hope of weaving scientific investigation into the fabric of their careers.

There are many facets to a scientific career, and it is impossible to expose trainees to all aspects equally. The resources available at different institutions vary greatly. There are currently two common models for research training within residencies. One is to assign residents a particular laboratory or investigator who will then assign a project. Typically, the resident simply learns the necessary technique in use by that investigator, generates some data, and writes up the project, functioning as a rather high-level technician. The strengths of such an approach are that the residents are virtually guaranteed a publication and/or presentation, and they get some experience with the chosen techniques, collecting and managing data, and working with scientists. An alternative model is to require the residents to identify their own research question, find a suitable mentor to help frame a testable research hypothesis, and design and execute an experiment. The strength of this approach is that it is a realistic simulation of a life in science. Most of the work involved in conducting research is in the conceptualization of the project. One can hire technicians and research assistants to "turn the crank" on a project, but only the principal investigator can do the intellectual work of asking the right question and designing the experiments to answer it. Unfortunately, this approach to research training engenders a great deal of frustration and often fails to yield a completed project. Whichever model is used, residents' training will be enriched by an opportunity to conduct a scientific investigation.

This chapter is about conducting clinical research. Conceptually, performing scientific investigation in the clinical domain is no different from the laboratory. There are the same requirements for sound research design, data collection, data analysis, and interpretation. In practice, however, there are some significant differences. Most importantly, clinical research is performed on human subjects. This imposes both practical and ethical considerations. Subject accrual, control groups, informed consent, and the generalizability of findings are but of few of the issues that arise. It is impossible to present a comprehensive course on clinical research in this brief chapter. There are many excellent books on the topic, some of which are listed in the Suggested Readings section at the end of this chapter, and the reader is strongly encouraged to seek such resources before embarking on a project. What follows here is an introduction to some of the important topics bearing on the conduct of clinical research.

WRITING A RESEARCH PROPOSAL

A written research proposal is an essential part of every research project. A typical research proposal is composed of an introduction that states the research topic, a list of specific aims or objectives of the project, a summary of the background and significance of the proposed work, a research plan detailing the method and materials, and a budget with justification of all proposed expenses. The proposal serves many functions. It will identify the research question and hypotheses. It will state the significance of the work. It is the first opportunity for the investigator to lay out the logical arguments supporting the research, enabling the person to visualize the project as a whole and in the context of whatever clinical questions and scientific studies have come before. It will codify details of the proposed methodology, necessary resources (financial, personnel, equipment, supplies, space, etc.), and time frame of the study. It provides a sort of road map for the collaborators on the project to be sure they are all in agreement about the structure, work assignments, and execution of the project. It is often used as the basis for a grant application to obtain funding for the project. It is typically used as a draft version of the introduction and methods section of any publication coming from the project. A well-written research proposal launches a project.

Research projects begin with a question. There are three criteria that a research project must meet: (1) the topic must hold the researcher's interest for the duration of the project, (2) the project must be doable with available resources (time, money, personnel, equipment, etc.), and (3) the project must be nontrivial. Therefore, identifying the research question is very important. Questions arising from the young researcher's own curiosity tend to be more motivating than those assigned by others. Often a clinical dilemma, interesting or poignant case, or unexpected finding will stimulate curiosity. This curiosity should lead to the literature to learn what is already known about the topic. Initially, such reading may be in review articles or textbook chapters but rapidly moves into primary source literature. Once the relevant clinical literature is exhausted, reading progresses into related basic science that bears on the topic. Eventually, one reaches a frontier where there is no scientific foundation for clinical practice, where clinical practice is supported only by tradition, empiricism, and folklore. This boundary is where biomedical science really takes place. The immediate research question then becomes obvious: it is the next thing that needs to be learned to build a foundation for clinical practice. Articulation of this question frames the research.

A research project cannot be based on a vague or general description of the investigator's curiosity. It must have very specific objectives that, once enumerated, guide the design of the proposed experiments. In "grant speak," these objectives are called "specific aims." A well-conceived project typically has anywhere from one to four specific aims. One can write a short overview in a few sentences stating the research question, background, and significance, then enumerate the specific aims. The statement, "We will see what happens when we give gentamicin eardrops to patients with external otitis" is not a specific aim. There are many different styles for writing specific aims. A particularly effective style is to state each aim in terms of the hypothesis to be tested and the method to be used for that part of the experiment. For example, "We will test the hypothesis that administration of gentamicin eardrops accelerates resolution of external otitis by comparing the results of serial ear canal cultures in a group of external otitis patients receiving gentamicin drops to a control group of external otitis patients receiving acetic acid drops." Anyone reading the proposal, then, has a short synopsis of exactly what the investigator is trying to accomplish and how.

The background and significance of the project may best be understood by the analogy of building a pyramid. Every fact discovered by scientific research is one brick in a pyramid. Within any giant pyramid, there are a multitude of smaller pyramids. One can think of the giant pyramid as the "big picture." For example, curing cancer is a giant pyramid that is still under construction. Embedded within it are countless other pyramids; there is one with discovery of oncogenes at its apex, another with *p53* at its apex, another with cell surface marker antigens at its apex, another with sensitivity to cisplatin at its apex, etc. If you are a pyramid builder, no matter how much you want to place a brick in a specific place, you cannot do so unless someone has laid the other bricks upon which yours will rest. You cannot suspend a brick in space. Thorough and complete description of all the previous research and clinical knowledge that lead to and support your proposed project is the background. It begins several layers deep and climbs to include all the work immediately contiguous to the current proposal. It is the small pyramid that has your proposed project as its apex. The surrounding larger or giant pyramid in which your project is embedded is the significance. Whether the background and significance are written in bullet form, outline form, or narrative, they provide a chain of logic that should convince the reader that the proposed project is the necessary next important brick to lay.

Once the research proposal has stated the research question, listed the specific aims, and described the background and significance, it must give details of the method and materials, or research plan, that will accomplish the stated objectives. The research plan should state the research design: retrospective versus

Outline of study protocol
Title
Research questions (*objectives*)
Significance (*background*)
Design:
Subjects
  Selection criteria
  Sampling design
Variables
  Predictor (independent)
  Outcome (dependent)
Statistical issues
  Hypotheses and analytic approach
  Sample size and power

Figure 13–1 *Research proposal outline (from Hulley and Cummings [1987]).*

prospective, cohort versus case control, longitudinal versus cross-sectional, etc. (see Research Design). It must include a description of research subjects, including inclusion and exclusion criteria that will qualify them for the study. There should be a clear statement of predictor (independent) and outcome (dependent) variables and how they will be measured. There must be a description of controls and a justification for their selection. The proposed sampling methods and measurements should be described, as well as a calculation of sample size necessary to achieve statistically meaningful results. Data analysis and statistical methods must be described. It is often helpful to include a section on anticipated problems or pitfalls and how those may be dealt with if encountered. It is worthwhile to be detailed and thorough in this part of the research proposal because it can help avoid methodological flaws that could ultimately undermine the validity of the results. **Fig. 13–1** is an outline of a research proposal. Completion of each section will form the basis of a complete and detailed written research proposal.

RESEARCH DESIGN

The design of a research study depends on the hypothesis to be tested and the resources available for realistically performing the study. There is an underlying assumption that findings in the study will be generalizable to the real world. The extent to which this is a valid assumption depends on the experimental design. If the population studied is a representative sample of the real-world population of interest, generalizability of results is enhanced. For example, if one wishes to assess the efficacy of a new antihypertensive drug in senior citizens, studying the drug's effect in volunteer medical students is of questionable benefit. Testing that same drug in attendees at the weekly bingo game at the local senior center should be better. However, if 90% of the bingo players are women, it may be risky to assume equivalent drug efficacy in men. The choice of experimental design seeks to balance reliability/generalizability of the study and cost in time and money.

The process begins with identification of the actual, real-world population of interest, the population to whom these results will ultimately be generalized. This enables the investigator to define inclusion criteria for suitable research subjects. Then the inclusion criteria are applied to an accessible target population that is a representative subset of the population of interest. Exclusion criteria eliminate those potential subjects who are unable to participate, who may provide bad data, or who are unethical to study. The remaining pool of candidates is either studied in its entirety or must be sampled. Sampling methods include consecutive samples, probability samples (e.g., random sampling), and nonprobability samples (e.g., convenience or judgmental).

In an experimental study, research subjects identified in the manner described here are assigned by the investigator to experimental and control groups who differ only by the nature of the experimental intervention. Assignment is done by a predetermined systematic means such as randomization. In a quasi-experimental study, the investigator is not able to make assignment into experimental and control groups. Usually this is a limitation imposed by the logistics of performing the study. For example, to compare the efficacy of two different treatments, one may need to compare subjects given treatment A at one institution to subjects given treatment B at another institution. The choice of which treatment is administered at each institution is not in the investigator's control. The two cohorts may differ in ways that undermine the validity of the study, but there is no practical alternative means of accomplishing the research.

There are many different designs, varying in complexity, reliability, and applicability. The inexperienced investigator is strongly advised to seek guidance from an expert in clinical research or consult one of the many excellent texts on the subject. It is essential that decisions of design, data collection, and data analysis be made before the project begins. Imagine you are a carpenter building a house and, once the walls are all finished, calling in an electrician and requesting installation of some lights and wall sockets. This is the situation faced when someone compiles a mass of data and then goes to the statistician for help with analysis. The experimental design, data collection method, type and amount of data collected, and statistical methods are all interrelated.

The following is an introduction to some of these most important factors

Research Design Categories

Research design can be broken into three general categories: experimental, quasi-experimental, and ex post facto (i.e., retrospective), in descending order of reliability.

In experimental design one group of subjects is exposed to an experimental intervention, and one or more other control groups are treated differently. The assignment of treatments is under the control of the investigator, and the assignment of subjects to the various groups is by some predetermined systematic means, such as randomization, to ensure that experimental and control groups differ only on the basis of the experimental intervention. This is the model routinely used in laboratory animal experimentation. It is the design most likely to yield unambiguous results or determine causality and least likely to be confounded by various forms of bias. Its use in clinical research is more restricted, but it is still the ideal for obtaining results of the greatest validity.

Quasi-experimental design, as the name implies, looks like experimental design but differs in that the investigator cannot assign subjects to an absolutely equivalent control group. For example, subjects may not be assigned randomly to two different treatments, they may vary in age, severity of their disease, temporal factors in duration of symptoms or time since treatment, site of treatment, or any host of other variables. Internal validity of a study, the likelihood that an observed effect of intervention is truly due to that intervention, is weaker in quasi-experimental design. It is the lack of experimental controls that introduces many threats to internal validity because there may be many variables that are not appreciated or controlled by the investigator. It is the job of the investigator to try and identify a population of patients to serve as approximate controls. The more closely the control group resembles the experimental group, the stronger the inference that can be drawn from the comparison. Cohort comparison and case-control studies are examples of quasi-experimental designs that can have good reliability.

Ex post facto design refers to analysis of groups after the fact. The investigator has no control of assignment to experimental groups. In other words, the investigator has no control of independent variables because they have already occurred. Sometimes this is the only practical type of research to answer certain questions, for example, in a disease so rare that it may take many years to accrue a significant number of patients for study. However, retrospective studies are also the weakest in reliability because of the loss of investigator control of random treatment assignment and control of variables. Retrospective studies are the most likely to lead to

Table 13–1 Hierarchy of reliability of research methods for ascertaining diagnostic or treatment effectiveness

Research method (in descending order of level of reliability)

True experiments (investigator controls both allocation to groups and determination of treatment)

Randomized concurrent controlled trial including crossover design with random order of treatment
Historical controls only in special case of certain diagnosis and known course of events
Randomized concurrent controlled trial with weakly randomized assignment or systematic assignment (odd/even, alternate appearance, etc.), including crossover design with systematic order of treatment
Nonrandomized concurrent controlled trial
Short-interval sequential trials within same institution or service
Controls from separate institutions or services with documented attention to coordination
Controls from separate institutions or services with poor or no attention to coordination

Nonexperimental methods

Cohort comparison studies
Historical controls
Nonexperimental ease control or case referent studies
Series of eases (all corners to a center over a time interval)
Large series of cases (consecutive)
Small series of cases (consecutive)
Isolated case reports with documentation of active surveillance
Isolated case reports (volunteer)
Case report

incorrect interpretation, especially by faulty conclusion of causality, a cause-and-effect relationship that can never be ascertained retrospectively. The greatest methodological flaw is to approach ex post facto data with no specific hypothesis or prediction in mind, but to just go “data mining,” looking for statistical associations. Such associations invariably show up but are uninterpretible. At best, these associations can be considered tantalizing findings that justify a prospective experimental or quasi-experimental study. **Table 13–1** (from Troidl et al. [1998]) gives a hierarchy of reliability in research methods in descending order.

DATA COLLECTION

Research data can be acquired in two manners: they can be collected by the investigators or their agents (observational data) or contributed by the research subjects (questionnaires and surveys). In large part, the choice of data collection method depends on what is being measured. Individual measurements should be sensitive, specific, appropriate, and objective, and they should detect differences over a range of values. Questionnaire and interview instruments should be clear, accurate, and reliable parameters that are assessable by validating an instrument in advance of using it for a research study.

The precision and accuracy of measures are paramount. Precision is defined as the consistency of repeated measures. It is a major determinant of sample size and statistical power. It is produced by random error introduced by variability in the observer, the measurement instrument, or the research subject. Trained observers using a manual of procedures developed for the project will generate more precise results; so will automated instruments and averaging results of repeated measures. Accuracy is defined as the degree to which a taken measurement actually reflects the true value of the variable. It is reduced by systematic error (bias) in the observer, instrument, or study subjects. Calibration, unobtrusive measures, and blinding improve accuracy.

Questionnaires and surveys can be self-administered or administered in a structured interview. In either case, when possible one should use existing instruments whose validity is already known. New instruments must be validated before use, a laborious process best done in collaboration with someone expert in questionnaire design. Although open-ended questions are useful for uncovering new or unanticipated information, closed-ended questions are more amenable to data analysis and are usually easier to answer. Generally, questions should be independent of each other, but they may be combined into summative or cumulative scores to gain an improved measure of abstract variables. The investigator should strive for completion of all questions in each questionnaire to avoid missing data that will complicate analysis and weaken the results.

Control and Validity

Research investigators seek to control three aspects of an experiment: the strength of the intervention, the equivalency of the experimental and control groups, and the equivalency of the study population and the real-world population to whom results may be generalized. These three levels of control are relatively straightforward for the bench scientist but can be extremely challenging for the clinical investigator. Strength of intervention may be the hardest to deal with. In a drug or radiation therapy trial, this is straightforward and relates to dosage. In other clinical interventions, such as teaching and training or surgery, standardization of the intervention across subjects may be impossible.

The relationship between the study group and the outside world is called external validity. Because one ultimately hopes to make results of the study applicable to the general population, the study population should be a representative subset of the real-world population of interest. Accurate definition of the sample frame, the potential pool of research subjects, is the first step in ensuring external validity. A further subset of the sample frame is taken as actual enrollees in the study. The sampling technique (see next section) by which subjects are chosen can also help ensure external validity by avoiding introduction of sample bias. Proper sample size must be calculated to determine how many subjects will be necessary for valid statistical analysis of results. Allocation of subjects to experimental and control groups should be random whenever possible. Nonrandom allocation introduces another potential bias that differentiates the experimental and control groups. Identification of sample frame, sampling method, calculation of sample size, and allocation of subjects are all under the control of the investigator in an experimental study. In a quasi-experimental design, some of these factors may not be under the investigator’s control. This threatens the external validity of the research. Other threats to external validity include interaction effect of tests or measure in the study that may alter subjects’ sensitivity to experimental intervention, interaction effects of selection bias and experimental intervention arising from known or unknown differences between experimental and control groups and their differential sensitivity to the experimental intervention, and multiple treatment interference in which effect of one treatment is not “washed out” before a next intervention is implemented.

Though not a complete list, these are some of the more common examples of threats to external validity.

The relationship between experimental and control groups within a study is called internal validity. Threats to internal validity include history, maturation, testing, instrumentation, statistical regression, selection or allocation bias, experimental mortality (dropout rate), and unanticipated interaction between any of these other factors. History refers to those factors that are going on in the world outside the confines of the study that may differentially affect experimental and control groups. Maturation refers to factors that may affect both experimental and control groups attributable to the duration of the study. Testing threatens internal validity when study subjects learn to anticipate the test or gradually improve their test taking by repeated experience. Instrumentation can be a source of trouble if calibration changes over the duration of the study or if technology introduces new measures before and after intervention. Statistical regression refers to inherent variability of any measured variable among study subjects. Subjects may have higher or lower scores unrelated to the experimental intervention. Nonrandom sampling or allocations are the most obvious source of bias and threat to internal validity because it reduces the equivalency of control and experimental groups. Experimental mortality is the loss of subjects to completion of the study. The rate of dropout and the cause of dropout may be different between control and experimental groups, introducing bias or incorrect interpretation. Finally, there can be interaction between these factors. For example, if control and experimental groups are allocated nonrandomly, the resultant cohorts may be differentially susceptible to maturation bias.

Sampling Methods

The sampling method is the means by which investigators select a representative subset of their target population for inclusion in a research study. The method is chosen to strike a balance between the "representativeness" of the study subjects and practical considerations such as time, expense, and availability of adequate numbers of subjects. Ideally, random selection is the best way to sample. In fact, sampling can be divided into probability and nonprobability methods. Probability methods are used when the chance that an element will be in the population is known. This enables a random selection technique to provide a representative sample. Nonprobability methods are used when the chance that an element in the population will be selected in the sample is unknown. The sample will not be representative of the population, and therefore research results can only be suggestive of statistical characteristics of the population.

There are four basic types of probability sampling: simple random, systematic, stratified random, and cluster sampling. Simple random sampling is applied to a homogeneous population in which every element in the population has an equal likelihood of being selected. The probability of each element may be different, but all elements are independent of each other. Simple random sampling is both simple and representative, but failure to accurately identify the entire population may introduce bias. Systematic sampling consists of selecting study subjects according to a predetermined sequence, for example, taking every fifth patient of the next 100 patients seen with the target disease. As long as the elements of interest are uniformly distributed through the sequence of patients, those chosen will be a representative random group. If, however, there is clustering of some element (e.g., all the older patients coming on the same bus from their nursing home), some elements may be over- or underrepresented in the study cohort. In stratified random sampling, the study population is subdivided into homogeneous groups based on known variables, such as age, economic status, tumor stage, and so on, to improve representation. This method works only if the researchers know which variables are necessary to achieve representativeness. Cluster sampling typically is used when the target population is infinite, potential subjects are widely scattered geographically, or there is no list identifying members of the population. Cluster sampling is often done in stages. However, as the number of sampling stages goes up, there are more chances for sampling error. Unappreciated differences between clusters or areas of sampling can introduce bias as well.

Nonprobability sampling methods include convenience sampling, judgment sampling, network sampling, and quota sampling. In the use of convenience sampling, subjects are selected by their accessibility. The validity of results will depend on how much the sample differs from the target population. Judgment sampling is when the researcher selects representatives of typical cases for study. The quality of the sample depends on the accuracy of the researcher in selecting representative cases. In network sampling a few subjects are chosen, and they, in turn, refer other subjects, who then refer a third tier. The networking continues until subject accrual is adequate. The technique is helpful when potential subjects may not readily make themselves known, but it has obvious bias because the subjects are not truly independent of each other. Quota sampling is similar to stratified random sampling. Bias from the differences between subgroups can be reduced by applying mathematical formulas to correct for underrepresentation. However, there is still little or no control of selection within groups.

Power and Sample Size

In research studies, power is defined as the probability of concluding there was a difference when in fact there was none. Sample size refers to the number of study subjects. One objective of study design is to enroll several subjects to permit valid statistical analysis of results. It is a number that can be calculated in advance of the initiation of the project based on the details of the test hypotheses and experimental design. For analytic studies, there are four steps to estimating sample size: (1) state the null and alternative hypotheses, specifying whether one or two-tailed; (2) select an appropriate statistical test based on the type of predictor and outcome variables; (3) estimate the effect size and its variability from pilot data or previous studies; and (4) specify appropriate values of α and β based on the importance of avoiding type 1 and type 2 errors. The actual sample size can then be looked up in statistical tables available for this purpose. For descriptive studies, the investigators must achieve a sample size that will provide a chosen confidence level and precision. The steps are (1) for a dichotomous variable, estimate the proportion of subjects with the variable of interest; for a continuous variable, estimate its standard deviation; (2) specify the desired precision (width) of the confidence interval; and (3) specify the confidence level (e.g., 95%). Sometimes the sample size is fixed or predetermined. If so, one can calculate backward to estimate the power or detectable effect size. The importance of addressing issues of sample size in advance cannot be overstated. It is this process that determines the feasibility of a project; if it is impossible to accrue enough subjects for statistical analysis, the project cannot be accomplished. In many cases, alterations in study design, such as choice of variables or statistical test, may enable performance of the research with a smaller sample size. Some common strategies include using continuous variables, more precise measurements, paired measurements, unequal group sizes, and more common outcomes. Collaboration with a statistician for sample size calculations and study design is essential.

DATA ANALYSIS AND STATISTICAL METHODS

Role of the Statistician

As already noted, statisticians play crucial roles in research design. They are the individuals best qualified to perform power and sample size calculations. They can help choose between different types of variables and measures based on objectives of generalizability and outcome validity.

All researchers should have a passing familiarity with the basics of biostatistics to discuss the issues intelligently with their consulting statistician. All physicians should have a passing familiarity with the basics of biostatistics so they can read the medical literature critically. Providing a solid primer on biostatistics is beyond the scope of this chapter. The objective here is only to introduce the reader to some basic biostatistics terminology and concepts.

Statistics provide a means of quantifying observations. The quantities can then be compared or manipulated to provide insights into those things measured. Statistics can be used either to describe a group (of people, of numbers, of illness, etc.) or to assess how well characteristics of a group can be generalized. These two applications are called descriptive statistics and inferential statistics. Statistics are performed on variables, where *variable* is defined as what is being observed or measured. Variables can be dependent or independent. The dependent variable is the outcome of interest. The independent variable is the intervention or what is being manipulated. The variables can be either discrete or continuous. Discrete variables can have only a limited set of values, such as gender, eye color, and number of offspring. Continuous variables, such as height and weight, fall along a continuum with units imposed by the sensitivity of the measuring technique. Discrete and continuous variables lend themselves to different types of statistical treatment.

Variables can also be grouped by type: nominal, ordinal, interval, and ratio (mnemonic = NOIR). Nominal variables are named categories, such as eye color, gender, and side of lesion. Ordinal variables are the same as nominal plus ordered categories (e.g., cancer stages I–IV). Interval variables are the same as ordinal plus equal intervals; that is, the difference between numbers is meaningful, but the ratios between them are not (e.g., intelligence quotient). Ratio variables are the same as interval plus meaningful zero; that is, the ratio between numbers is meaningful (e.g., weight). Once again, the different types of variables are amenable to analysis using different statistical methods. Parametric statistics are used for interval or ratio dependent variables. Nonparametric statistics are used for nominal- or ordinal dependent variables.

Descriptive Statistics

Descriptive statistics are concerned with the presentation, organization, and summarization of data. The first important description of data is the central tendency.

The best-known calculations are mean, median, and mode. The second important description is dispersion: how closely the data cluster around the measure of the central tendency. Dispersion can be reported as the range; that is, the difference between the highest and lowest values of the variable. A more informative description is standard deviation. Standard deviation (SD) defines how closely individual scores cluster around their mean. $SD = \sqrt{\Sigma x^2/N}$, where x is the deviation of an individual measure from the mean, Σx^2 is the sum of the squares of all values of x in the sample, and N is the sample size. Standard error of the mean (SE) describes how close mean scores from repeated samples will be to the true (population) mean. $SE = SD/\sqrt{N}$, where SD is the standard deviation and N is the sample size. Both SD and SE assume a normal distribution, the situation in which all values are evenly distributed above and below the central tendency. If the distribution is asymmetric about the central tendency, the asymmetry can be described by skew and kurtosis, and requires different means of expressing dispersion.

Inferential Statistics

Inferential statistics allow one to generalize from sample data to a larger group of subjects. One can define statistical inference as the determination of the probability (or likelihood) that a conclusion based on analysis from a sample is true. All statistical tests are based on the signal-to-noise ratio, where signal is the important relationship and noise is a measure of individual variation. There are four possible outcomes of signal-to-noise: (1) detecting a signal where there was none (false-positive), (2) detecting a signal when there was one (true-positive), (3) detecting no signal when it actually was present (false-negative), and (4) detecting no signal when there was no signal (true-negative). Statisticians define two terms, a (α) and b (β), to talk about erroneous measures. Alpha is the probability of concluding that the sample came from a different population (i.e., a significant difference exists) when in fact it did not (making a type 1 error). Beta is the probability of concluding that no difference existed when in fact it did (making a type 2 error). The power of the statistic, mentioned earlier, equals $1 - \beta$. The relationship between α and β is shown in **Table 13–2**.

When the statistical test can detect any difference between groups, regardless of direction, it is a two-tailed test. A one-tailed test specifies the direction of difference in advance. Confidence intervals define the range within which the true mean of a population falls; for example, 95% confidence intervals ($\pm$2 SE) have a 95% chance of containing the true mean. If the 95% confidence intervals of two populations overlap, then the difference between them is not significant at the .05 level. Statistical significance is a precondition for a consideration of clinical significance but says nothing about the actual magnitude of the effect. For example, consider a new drug that has only one tenth the risk of ototoxicity as the drug in current use. The difference in ototoxicity between the two drugs is highly statistically significant. However, if the risk of ototoxicity of the standard drug is a low 0.01%, the 0.001% risk of the new drug is not a clinically significant improvement.

Table 13–2 Relationship between α and β

	Truth	
Called	**No Difference**	**Difference**
No difference	(1-α)	β
Difference	α	(1-β)

Parametric statistical tests of significance are applied in cases when outcome variables are interval or ratio. The best known is the t-test, a method of comparing the means of two groups. It is based on the ratio of the difference between groups to the standard error of the difference. The unpaired t-test compares the means of two independent samples; the paired t-test compares two paired observations on the same individual or matched individuals. The t-test is not appropriate when there are more than two groups or when individuals in one group are matched to individuals in another. Analysis of variance (ANOVA) is used to compare among many means. It is used when the independent variable is nominal and the dependent variable(s) is/are interval or ratio. A one-way ANOVA deals with a single nominal independent variable; a factorial ANOVA deals with multiple different factors in many different configurations. Regression analysis is used in the situation in which there is one measured outcome/dependent variable and one or more measured independent variables, when both dependent and independent variables are interval or ratio. The Pearson correlation and the multiple correlation coefficients describe the strength of the relationship between variables. Analysis of covariance (ANCOVA) combines regression and ANOVA when there is one measured dependent variable and independent variables can be both categorical and measured.

Nonparametric statistics are applied when the outcome/dependent variables are nominal or ordinal. The chi-square, binomial test, and Fisher exact test are all for nominal data and independent samples. The McNemar chi-square test can be used for related samples. When data are ordinal and samples are independent, Mann-Whitney *U*, median, Kruskal-Wallis, and Kolmogorov-Smirnov tests may be applicable. There are several nonparametric measures of association equivalent to the correlation coefficient. These include contingency coefficient, phi coefficient, and Cohen's kappa coefficient for nominal data, and Spearman's rho, Kendall's tau, and Kendall's *W* for ordinal data. There are three advanced nonparametric techniques for handling designs where the dependent variable involves frequencies within categories, with more than one independent variable. The Mantel-Haenzel chi-square deals with two independent factors. Logistic regression and log-linear analysis can manage any number of independent variables. Logistic regression treats all independent variables as measured data like multiple regressions. Log-linear analysis handles the case of multiple categorical variables and estimates effects and interactions, analogous to factorial ANOVA.

In addition to this brief list of parametric and nonparametric statistical tools, there are a host of others, as well as multivariate techniques. There are many software packages available today to enable anyone to perform these statistical tests on a desktop computer. However, the software cannot assure the quality of data nor prevent you from using the wrong statistic. Only the statistician can do that.

Interpreting and Reporting Research Findings

As pointed out in the previous discussion, clinical research studies should be as rigorous in their statistical design as basic science research experiments. This point is especially critical when examining relatively rare disorders where large populations may be necessary to achieve statistical significance. The process of interpreting data published in the medical literature should essentially involve reverse engineering the project. Initially, one should determine if the results actually addressed the key research question asked. Next, one should ask if the reported aims of the paper were achieved by the research design and if the statistics chosen to prove the null hypothesis were adequate. This same process should be followed when reporting your own research findings. In especially difficult research situations, the results of the studies should be reported in enough detail so that they are suitable for later meta-analysis by other groups of investigators. If the research question is properly posed and the research project well designed, results worth publishing or results leading to a follow-up study should emerge. If these points are not considered, one's time expenditure and utilization of resources will result in a project that neither addresses nor answers any significant questions.

Suggested Readings

Hulley SB, Cummings SR. Designing Clinical Research: An Epidemiologic Approach. Baltimore: Lippincott, Williams & Wilkins; 1987

Norman GR, Striener DL. Biostatistics: The Bare Essentials. St. Louis: Mosby Yearbook; 1994

Okolo EN, ed. Health Research Design and Methodology. Boca Raton, FL: CRC Press; 1990

Troidl H, McKneally MF, Mulder DS, Wechsler AS, McPeek B, Spitzer WO, eds. Surgical Research: Basic Principles and Clinical Practice. 3rd ed. New York: Springer-Verlag; 1998

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. A type 2 error refers to
   A. An error in calculations made by data entry
   B. Mistakes made through improper study design
   C. The probability of erroneously identifying that there is no difference between two groups
   D. Incorrectly detecting a false-positive result
2. Parametric statistical tests include
   A. Chi-square test
   B. Spearman's rho
   C. ANOVA
   D. Kendall's *W*
3. Precision refers to
   A. How often a given question is answered correctly
   B. The consistency of repeated measures
   C. The degree to which a measure is a correct metric for the question being asked
   D. The independence of variables

Chapter 14

BASIC PRINCIPLES AND CURRENT APPLICATIONS OF LASERS IN HEAD AND NECK SURGERY

DANIEL B. KURILOFF

The first visible light laser (a synthetic ruby crystal) was developed in 1960 by Theodore Maiman at the Hughes Research Laboratories (in Malibu, CA). It produced visible red light lasting only a few microseconds. Industrial applications for laser energy rapidly followed with an ensuing explosion in technological refinements. In addition to its use as a laboratory instrument, the laser has become an integral part of the digital age (optical scanners, holographic imaging, compact audio and video discs, laser pointers, pulse oximetry, telecommunications, industrial welding and drilling, etc.).

It was quickly recognized that the unique properties of laser energy would allow it to be harnessed for use as an ideal surgical tool, providing unparalleled user control for bloodless, inherently sterile tissue incision and ablation. Furthermore, the ability to deliver the beam to previously inaccessible regions of the body with precision, often at some distance, and the ability to incise muscle without electrical stimulation have contributed to its widespread use in surgery.

Laser photocoagulation of retinal vessels was performed by the mid-1960s, only a few years after its initial discovery as a laboratory instrument. In 1965, the carbon dioxide (CO_2) laser was developed, and 3 years later an optical articulating arm was introduced that allowed precise delivery of the beam to the upper airway. Use of the laser in ophthalmology and otolaryngology quickly spread to other medical fields, along with a several-fold increase in site-specific applications throughout medicine. Fueled by a rapidly growing industry for medical applications, laser technology advanced with the introduction of a plethora of new lasers—argon, potassium titanyl phosphate (KTP/532), CO_2, neodymium:yttrium-aluminum-garnet (Nd:YAG), erbium:YAG, holmium:YAG, thulium-holmium-chromium (THC):YAG, Q-switched YAG, Q-switched ruby, candela, alexandrite, excimer, gold vapor, copper vapor, mercury vapor, pulsed dye, tunable dye, holmium, diode, free electron lasers, and others—each with unique properties best suited for specific applications. Novel delivery systems (optical fibers, wave guides, contact tips) and new beam distribution technology (swift lase, silk touch, ultra pulse, flash scan, etc.) have evolved to improve controlled surface tissue ablation (e.g., laser skin resurfacing and epilation) and permit minimally invasive endoscopic application. Specialized laser instrumentation has also evolved to improve smoke evacuation and visibility and to protect adjacent nontarget structures. The use of laser light in conjunction with photosensitizing agents promises to become a major primary and adjunctive treatment for accessible superficial epithelial tumors.

In the last decade, the use of laser technology has permeated the ambulatory surgery environment and more recently has become a standard piece of equipment in the office setting. Solid-state lasers have become diminutive in size, facilitating portability and creating a small office "footprint." The hemostatic properties of the laser are especially advantageous in upper aerodigestive tract surgery, which in the past, required endotracheal intubation to protect the airway from bleeding and obstruction. "Bloodless" surgery in this region of the head and neck can now be performed safely on an awake patient in an ambulatory setting. With the popularization of laser-assisted office-based procedures such as laser-assisted uvulopalatoplasty (LAUP), laser-assisted intranasal surgery (LAST), laser-assisted endoscopic laryngeal surgery (LAELS), laser-assisted myringotomy (CTOLAM), and laser-assisted skin resurfacing and epilation, the number of physicians using laser technology and the number of procedures performed are expected to dramatically increase.

LASER PHYSICS AND TISSUE INTERACTION

A detailed discussion of laser biophysics is beyond the scope of this chapter, and the reader should consult other texts for greater technical depth. The word *laser* stands for light amplification by stimulated emission of radiation. The generation of laser energy involves two principles of quantum mechanics and atomic theory first postulated by Albert Einstein: spontaneous emission and stimulated emission. Electrons, which orbit the nucleus of an atom, exist in a low energy or ground state. An external energy source will excite atoms within the lasing medium and force electrons from the ground level to a higher energy orbit. This quickly decays back to its ground energy state, spontaneously emitting a discrete amount of light energy called a photon. Stimulated emission is said to occur when an atom, already in an excited state, absorbs an additional photon of energy from the same type of atom and quickly decays back to the ground state. When this occurs, two photons of energy with identical wavelengths and phase are emitted. The atoms of a laser exist within the lasing medium (which may be any of three states of matter) encased within an optical, resonant chamber with parallel, inward-facing end mirrors. On one end of the chamber, the mirror is partially reflective, which permits the energy to escape from the aperture of the laser tube.

To begin the process of laser emission, an external source of energy, or "pump," is first used to excite the atoms of the lasing medium. This may be in the form of a flash lamp, electric arc, or another laser, as in the case of the YAG laser. The internal energy created by spontaneous emissions contributes to a cascade of stimulated emissions when incident to other atoms within the lasing medium, thus creating a form of energy amplification. Energy released within the laser tube travels in random directions and may be reflected between the two end mirrors, further amplifying the process (**Fig. 14–1**). When more than one half of the atoms are in the excited state than in the ground state, a population inversion is said to occur, and laser energy is emitted.

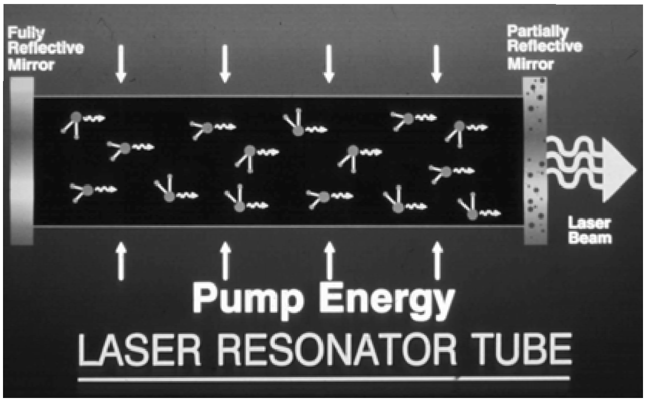


Figure 14–1 *A laser resonator tube, with a partially reflective end mirror, allowing escape of laser energy from the aperture after the lasing medium has undergone stimulation by the pump energy source and an atomic population inversion is said to have occurred.*

The energy released from the end of the laser tube has unique properties that distinguish it from ordinary light, which is disorganized or scattered in all directions from its source and composed of many wavelengths. Incident, reflected, and emitted light waves traveling back and forth between the mirrors of the laser chamber will eventually resonate in phase-producing coherent light (all waves in phase) (**Figs. 14–2 A,B**). The energy is monochromatic (one wavelength) based on the content of the lasing medium. Laser energy may be in the visible light spectrum, as in the case of the argon laser, or invisible, as produced from the CO_2 gas laser. Light that escapes from the aperture through the partially reflective end mirror will also become collimated, so that all light waves travel parallel to the source. With the aid of a lens, the parallel light from a laser can be focused down to the smallest possible spot size, a diffraction-limited spot. White light from a tungsten filament will focus to an inverted image of the filament but never to as small a spot. These properties (coherence, monochromicity, and collimation) allow the energy to be optically focused down to a small spot with little beam divergence or attenuation, creating a very intense beam of energy. The monochromicity allows selective absorption of the energy by different components of the tissue and is an important feature that permits photodynamic therapy, as described later.

Table 14–1 Absorptive heating by laser energy

Temperature	Visual Change	Biological Effect
≥100°C	Smoke plume	Vaporization, carbonization
90–100°C	Puckering	Drying
65–90°C	White/gray	Denaturization
60–65°C	Blanching	Coagulation
37–60°C	None	Warming, welding

These unique properties of laser energy allow it to be concentrated into a small target area from a distance and its biological effect controlled by varying the energy density and rate of delivery of the beam. The mechanisms of laser–tissue interaction involve photochemical, photothermal, and photomechanical/acoustic effects. The ultimate biological response is a complex summation of these effects on cellular and subcellular tissue components, the postsurgical inflammatory response, and long-term healing. The immediate biological response to a specific energy density will cause a physical effect, which ranges from warming and blanching to complete vaporization (see **Table 14–1**). The spread of thermal

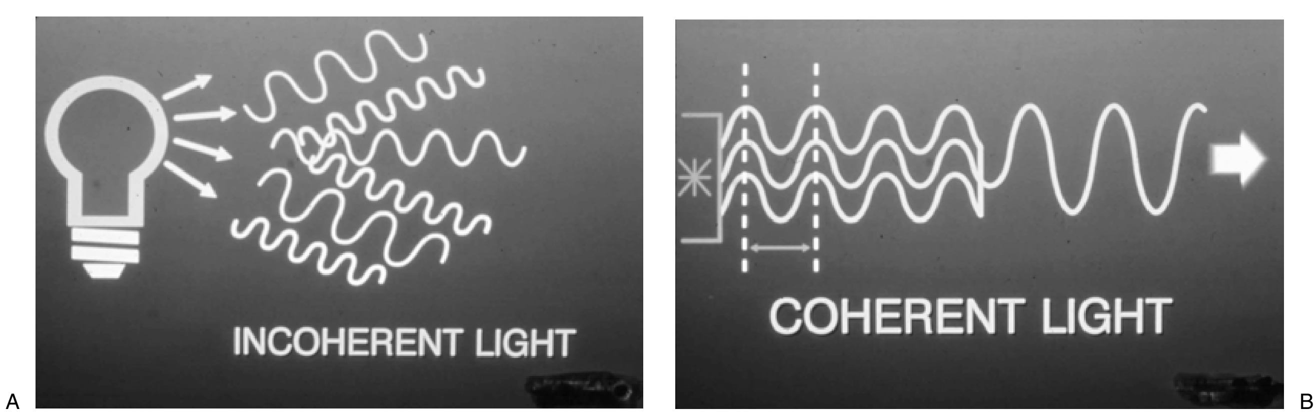


Figure 14–2 ***(A)*** *Disorganized or incoherent light waves from an incandescent lightbulb. Here, waves are emitted in all directions at random, out of phase, with mixed, broad-spectrum wavelengths.* ***(B)*** *An illustration of organized or coherent light waves from a laser tube. Here, the light waves are shown in phase with each other, all of the same wavelength, and parallel or collimated. The phase matching of the light generated by the laser further intensifies the energy emitted at the end of the tube.*

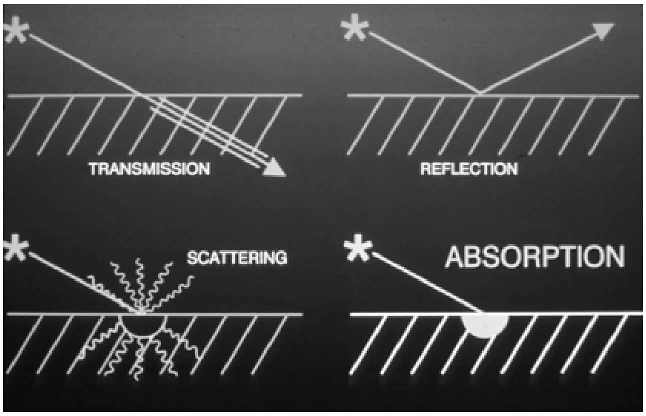


Figure 14–3 *Four types of laser light–tissue interactions. Both scattering and absorption will generate heat, which ultimately achieves tissue vaporization and denaturation. Transmission is essential when the target area is deep to the surface.*

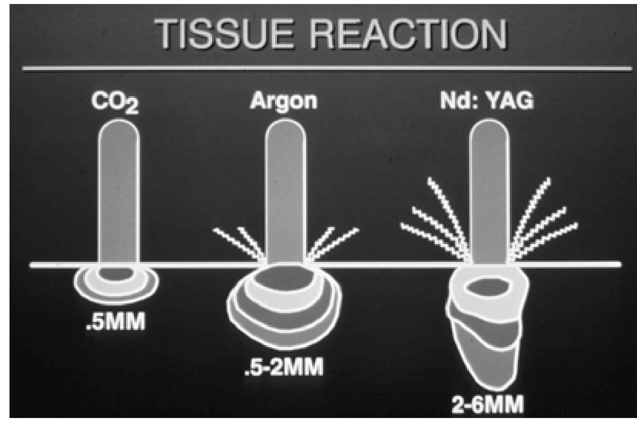


Figure 14–4 *Different laser wavelengths at a given power and spot size will penetrate to different tissue levels with varying degrees of thermal spread and destruction.*

energy through the tissue depends on the depth of laser penetration, the tissue's thermal conductivity, and the dynamic heat sink effect of the local microcirculation over the time of application. The depth of tissue effect also depends on the way in which the specific wavelength of energy is reflected, transmitted, absorbed, or scattered at the tissue interface, which in turn is dependent on specific tissue composition (e.g., pigmentation, water content, and vascularity)(**Fig. 14–3**). Reflection prevents penetration and creates safety concerns. Scatter will limit beam penetration and promote thermal spread. Transmission is critical when the target area is not located at the tissue surface (first-strike zone). For example, the ability of laser energy to transmit through the cornea, lens, and vitreous-filled eye chamber to reach the retina is essential for its ophthalmologic application. Scatter will limit beam penetration and promote thermal spread.

THE LASER AS A SURGICAL TOOL

The specific wavelength of laser energy will determine both its precision as a scalpel and its hemostatic properties. When used as a scalpel, the laser relies on its ability to cut tissue precisely with little thermal spread and tissue injury. In this case, laser energy must be intense, sharply focused, and absorbed almost entirely at the surface. The CO_2 laser is an ideal surgical scalpel because of its high absorption by tissue water content. It is therefore a surface laser that vaporizes superficially and allows a what-you-see-is-what-you-get type of tissue interaction similar to a cold steel scalpel or the electrocautery. It is less hemostatic, however, especially when encountering blood vessels that exceed the diameter of the laser spot size. At high power settings, the CO_2 laser primarily will vaporize tissue with a great deal of precision and little lateral thermal injury and char. In contrast, the Nd:YAG laser energy will transmit through water with little attenuation, being more selectively absorbed by pigmented tissue. The Nd:YAG laser tends to penetrate more deeply into tissue planes, resulting in a field of thermal injury much deeper and less visible than that achieved by the CO_2 laser (**Fig. 14–4**) The Nd:YAG laser is therefore a much less precise tool for use as a surgical scalpel but is one of the most hemostatic of all lasers. However, tissue necrosis can occur well beyond what is initially visible and therefore requires considerable training and experience for safe use. Similarly, the visible light lasers, the argon and KTP lasers, are more selectively absorbed by hemoglobin and pigmented tissues. Their precision as a surgical scalpel and hemostatic properties fall somewhere between the Nd:YAG and CO_2 laser energies (see **Table 14–2**).

POWER

In addition to the physical properties of the specific laser energy, tissue interaction is user controlled by adjusting the power, spot size, and exposure time of

TABLE 14–2 RELATIVE TISSUE EFFECTS OF DIFFERENT LASERS

	Precision	Hemostasis
CO_2	****	*
Argon/KTP	**	**
YAG	*	****

CO_2, carbon dioxide; KTP, potassium-titanyl-phosphate; YAG, yttrium-aluminum-garnet, increasing (*) asterisks > the effect.

the laser beam. Power is the least frequently altered parameter during a treatment session. In general, the highest safe power setting should be used when the laser is needed as a scalpel or for tissue ablation. At high power settings, most of the tissue ideally is vaporized, leaving a clean incision with minimal char and thermal injury. However, for laser welding (e.g., microvascular anastomoses and microflap, vocal fold phonosurgery), very low power settings may be used (milliwatts) to allow photocoagulation and denaturization as the predominant tissue effect.

Spot Size and Power Density

Because all of the laser energy is concentrated in the area of beam contact, altering the spot size will spread the same amount of energy at a given power setting (watts) over a greater or smaller surface area (cm^2) and reduce or increase the power density or irradiance ($watts/cm^2$). Irradiance equals power in the focal spot divided by area of the focal spot. Surface area and irradiance vary with the square of the beam diameter. Therefore, if the spot diameter is doubled, the irradiance will decrease to one fourth at a given power setting. If the spot diameter is cut in half, the irradiance will increase fourfold. This change in power density can be accomplished by focusing the beam with lenses and working within the focal point of the lens for greatest power density (smallest spot size) or defocusing the beam to reduce power density (larger spot size). In summary, the irradiance or power density can easily be altered in an exponential fashion by either changing the focal length of the lens or changing the working distance from laser to tissue plane (**Fig. 14–5**).

Treatment Time and Fluence

Although the total amount of energy delivered to the tissue and power density, or irradiance, is important, the rate of delivery may be the most crucial user-defined variable in determining the laser's ultimate biological effect. With a given spot size and power setting, the surgeon can vary the exposure time to achieve the desired result. The rate of energy delivery by the laser is called fluence. Fluence equals power density multiplied by time and is measured in watts ($W \times s/cm^2$) or joules (J/cm^2). For any given total amount of energy delivered per application, the rate of energy delivery will have the greatest biological effect. Control over fluence is, in actual practice, the most difficult concept to master and requires training, supervision, and experience. For example, a delivery of 100 J of energy can be achieved using either 25 W/cm^2 for 4 seconds or 100 W/cm^2 for 1 second, with a major difference in biological effect (**Fig. 14–5**). The majority of adverse tissue effects with laser technology occur because the inexperienced surgeon will tend to use an insufficient power setting over an excessively long exposure time. This tends to create a greater degree of thermal injury for the same degree of tissue ablation achieved at higher power settings over briefer exposure times. This results in excessive char, colateral tissue destruction, and subsequent scar tissue formation. Therefore, when using the laser to cut or ablate tissue, the surgeon should always use the highest level of fluence that he or she is comfortable with (**Fig. 14–6**). The fluence delivered to tissue also can be modified by using the laser in a pulsed delivery mode rather than using it in a continuous mode.

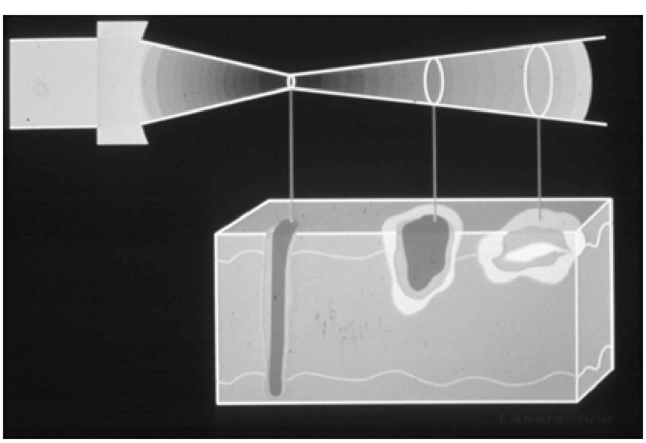


Figure 14–5 *Changes in power density, depth of tissue penetration, and lateral thermal spread with distance from the focal point as well as spot size.*

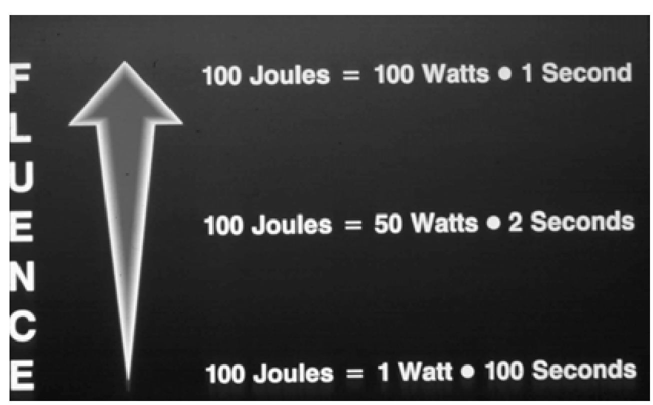


Figure 14–6 *The increase in fluence achieved by delivering the same total energy from the laser over a shorter time frame. This technique will optimize tissue vaporization with little lateral thermal tissue injury and is one of the most important techniques to master when using the laser as a surgical tool.*

Pulsed Delivery

The more commonly used lasers in the head and neck—argon, KTP, Nd:YAG, and CO_2—operate in the continuous wave mode. The novice will often use a pulsed delivery mode in the order of seconds to feel more in control of the energy delivery. Short-pulse sequences or super-pulsed modes will often allow higher energy delivery with less thermal injury by using the heat sink effect of the surrounding tissue and blood flow during the interpulse intervals. Some lasers are pulsed by rotating or removing one of the end mirrors. Q-switched lasers rely on this technique to produce bursts of laser emissions in nanosecond pulses. Q-switched pulsed lasers are most useful in achieving fine control over surface ablation, an important attribute for laser skin resurfacing.

LASER TRANSMISSION AND INSTRUMENTATION

Optical Fibers

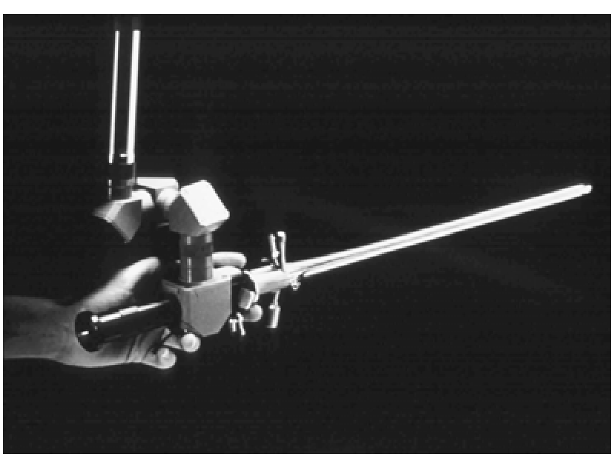
The energy emitted from the laser tube may interact with its target directly or more commonly is focused with lenses and transmitted through optically reflective wave guides, articulating arms, or fiberoptic cables (**Fig. 14–7A,B**). Not all laser wavelengths can be transmitted through true optical fibers, where there is total internal reflection and close to 100% transmission (with little or no heat generation). A wave guide is a type of flexible optical cable with a highly reflective, internal metallic coating. It differs from true fiberoptic light cables in a lower efficiency of light transmission and higher heat generation, often requiring higher power outputs and air cooling with miniature pumps built into the laser. For example, a cost-effective fiberoptic cable is not yet clinically available for the CO_2 laser. To reach areas within the nasal cavity, a wave guide is used instead. The laser energy also can be focused to a smaller, more intense microspot through the use of lenses within handpieces or the operating microscope.

Flash Scanners

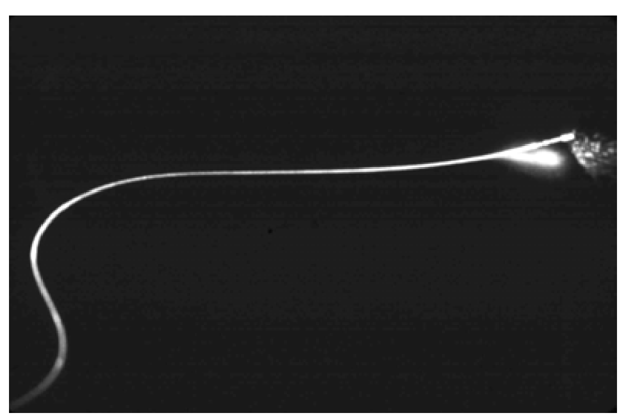
To spread the laser energy more uniformly over a target area, the flash scanner was developed. This is essentially a moving laser beam within a defined area created by two nearly parallel, rapidly rotating mirrors. With this technology, laser energy may be spread uniformly over an area with a diameter of 3 mm in 1 millisecond. Similarly, the laser energy may be pulsed rapidly over a defined surface area achieving a similar effect (see later discussion on skin resurfacing). Lasers merged with this technology are capable of precise surface ablation with little or no char at depths as little as 0.15 mm and is ideally suited for epithelial resurfacing (tattoo removal, wrinkle removal, lip vermilionectomy for leukoplakia, surface ablation of tonsillar crypts, etc.).

Laser Instrumentation

Laser hits from a reflected beam can cause both skin burns and eye damage. Metallic instruments that appear dull in visible light may in fact act like a mirror when exposed to the far infrared wavelength of the CO_2 laser. Reflected, misdirected laser hits can be avoided by using



A



B

Figure 14–7 (A) *A laser bronchoscope. The laser beam is reflected through a prism, and the spot position within the lumen of the scope is adjusted with a pivoting mirror attached to a joystick.* ***(B)*** *Laser energy (argon) may be transmitted with little energy losses through a fiberoptic light cable. Not all laser wavelengths can be transmitted in this way and require less efficient wave guides, which achieve less efficient energy transmission and generate more heat within the carrier.*

instrumentation that has a low specular or direct reflectance. Laser energy striking this type of instrument will produce a large diffuse or scattered and misdirected laser beam, with little risk of tissue injury. Attempts to blacken or ebonize and roughen the surfaces of metallic instruments will afford some but not complete protection. Ebonized instruments often lose their black coating after several cycles of sterilization and may have no real-life protective benefit over nonebonized instruments.

Some laser handpieces used within the oral cavity have a backstop to prevent laser hits beyond the target once vaporization is complete. It is essential that continued, prolonged laser exposure of the backstop be avoided because this may cause excessive heating and potentially burn the patient. Instruments should be lightweight to facilitate unencumbered movement, especially when attached to smoke evacuator tubing (see below).

LASER SAFETY CONTROL MEASURES

Education

The potential for disaster and the likelihood of litigation should not be underestimated when a medical laser is used by either inexperienced or careless individuals in an office setting or in an operating room. It is essential that every member of the surgical team, including the patient, be educated and informed. The patient should understand the specific benefits of the laser over other existing technology, and it should not be used unless those benefits have been well documented through clinical trials and peer review. The patient also must understand the unique risks associated with the use of laser energy and the potential for injury away from the operative field. This information should be included in the process of informed consent.

Safety Guidelines and Credentialing

In 1988, safety standards for laser use in medicine were first outlined in an American National Standards Institute (ANSI) report titled "Laser Safety in the Health Care Environment." The standard itself outlines specific procedural and administrative controls necessary to ensure the safety of patients and health care professionals working with lasers and intended as a guide to aid the manufacturer, the consumer, and the general public. Although the ANSI report has defined the current standard of care for the safe use of lasers in the health care environment, compliance in general is voluntary unless required by a specific organization (e.g., hospital or ambulatory facility).

Administrative controls include the establishment of a laser safety committee and appointment of a laser safety officer. The laser safety committee generally consists of a multidisciplinary group, including physicians, nurses, biomedical engineers, and hospital administrators, who meet on a regular basis to establish and enforce adequate protective measures against laser-induced injury and provide in-service training for the management of laser-induced catastrophes (e.g., fires, burns, and explosions). Credentialing procedures must be set up and monitored by the committee for physicians, anesthesiology staff, nurses, technical support staff, and other health care personnel. Educational programs of at least 16 to 20 hours in duration and with 50% of the time dedicated to hands-on experience have been suggested as a minimum requirement for credentialing. The program must include exposure to all of the specific wavelengths of energy that will be used, as well as comprehensive training in all aspects of laser biophysics and safety. As new wavelengths of laser energy are introduced, specific additional hands-on training would also be required.

The laser safety officer must ensure continuing surveillance and enforcement of safety regulations and credentialing requirements. Any laser-related complications or problems are documented and reported directly to the committee for review and appropriate corrective action. New laser installations require inspection by the laser safety officer, in conjunction with the manufacturer and/or biomedical engineer, and should include assessment of other special considerations such as electrical hazards, fire and explosion hazards, eye exposure risks, laser smoke evacuation, and the flammability of anesthetic agents and drapes.

Indeed, the incidence of serious laser-related injuries at institutions adhering to the ANSI guidelines remains astonishingly low despite an increase in laser use across surgical subspecialties. Accidents occurring during laser surgery are almost always related to breaches in safety protocol and surgical accidents. Medical assistants or nursing staff in the office must have already received laser credentials and hands-on training in laser safety, or they must attend courses for certification prior to assisting physicians in performing laser procedures. All personnel should have in-service training from the manufacturer of the specific laser(s) being used in the office.

General Safety Considerations in the Office or Operating Room

The treatment or operating room must be of adequate size to allow adequate dissipation of heat from the laser unit and permit enough room around the patient to allow unencumbered movement of the laser articulating arm or laser fiber. Specific electrical and plumbing requirements must meet local building codes and be in place prior to installation of a laser in the office. This may require consultation with an engineer or electrician to ensure the presence of adequate circuit breakers and fire protection. Smoke detectors and fire extinguishers should be present at all times and regularly tested for proper function.

In the event of laser malfunction, traditional surgical instruments and hemostatic capabilities must be readily available. An electrocautery unit should always be present as well as silver nitrate to control minor bleeding. To avoid any potential fire hazard, all flammable liquids (e.g., ethyl chloride, alcohol, and acetone) should be removed from the treatment room when using the laser. Nonflammable drapes and water-saturated towels should be used to protect both the patient and anesthesia circuit tubing when appropriate. In the event of an endotracheal tube fire, a bronchoscope should be readily available both to reestablish the airway after immediate removal of the endotracheal tube and to inspect the endobronchial tree for thermal and smoke injury. All basic life support and emergency equipment must also be available.

Because the laser is capable of causing combustion within the patient's airway and in the operating room, special precautions must be taken to avoid a fire hazard. Also, because most patients undergoing upper airway surgery with the laser require general anesthesia and the delivery of oxygen through an endotracheal tube, the least flammable mixture of inhalational agents should be used whenever possible. Mixtures of helium, nitrogen, or room air plus oxygen are often used to minimize the oxidizing effects of pure oxygen. Combinations of nitrous oxide and oxygen are particularly dangerous and should never be used because a "blowtorch effect" could cause serious injury to the tracheobronchial tree (**Fig. 14–8**). The risk of an airway fire may be minimized by the use of nonflammable laser-safe endotracheal tubes. Metal and coated tubes are available, but all have their advantages and disadvantages, and none are totally without risk. In addition, the use of double-cuffed endotracheal tubes filled with saline both prevent loss of airway protection due to inadvertent puncture by the laser and will help extinguish an airway fire. The proximal cuff is often filled with dilute methylene blue dye to help warn of a cuff puncture. Protecting the endotracheal tube from stray laser hits is essential. The use of saline-saturated cottonoids, laser suction platforms, and nonreflective, ebonized instrumentation are all essential safety precautions. Eliminating the use of an endotracheal tube during surgery by jet ventilation or apneic technique with intermittent reintubation is the safest method but requires the participation and cooperation of an experienced anesthesiologist.

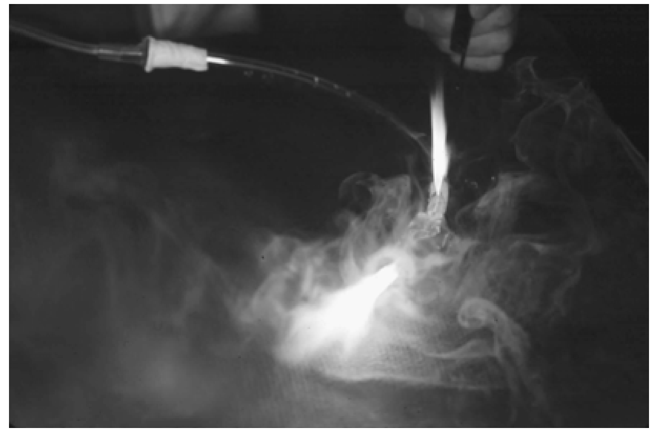


Figure 14–8 *A demonstration of the "blowtorch effect" resulting when flammable anesthetic gas mixtures are ignited by the laser.*

Laser Quality Control and Lockout Features

The U.S. Food and Drug Administration (FDA) heavily regulates the development and introduction of medical devices for use in the health care industry. All lasers manufactured or imported into the United States are regulated by and must conform to all safety regulations established by the Center for Devices and Radiological Health, a division of the FDA. These regulations classify each laser by power and wavelength and specify its mechanical and electrical requirements. Laser manufacturers are required to classify their laser by one of four major hazard categories, depending on the type and power output. Each classification has specific operational safety guidelines. All lasers must have an identifying decal stating the specific class and the need for eye protection. The majority of medical lasers, with the exception of low-power diode lasers and lasers used specifically for alignment purposes [e.g., the helium-neon (HeNe) laser], fall into class IV. Class IV lasers are hazardous to view either directly or indirectly, will cause skin burns, and are a fire hazard. Approved laser safety glasses of the appropriate spectral range must be worn at all times when the laser is turned on and in the treatment mode (see below).

All lasers must have a key switch interlock or lockout feature to ensure that only qualified individuals use them. Most lasers will have a key-actuated master control switch. The key is removable, and the laser cannot be operated when the key is removed. The laser should always be tested for proper functioning, and in the case of the CO_2 laser, the aiming beam and treatment beams should be in reasonable alignment and can be tested on a wet tongue depressor. If there is any doubt about the proper functioning of the laser, either the procedure should be rescheduled or more traditional surgical techniques should be used until it can be professionally serviced. Medical lasers should be serviced by their manufacturer on at least a biannual basis to ensure proper output and alignment.

Laser Warning Signs and Blackout Shades

A focused laser beam can traverse a room with very little attenuation and cause significant eye injury. It is essential that windows have blackout shades and entrance and/or exit doors be closed at all times to protect personnel outside the treatment or operating room. The ANSI-approved laser warning sign indicating the laser's specific class, power, and wavelength should be hung in clear view outside the room, and ideally a "laser in use" status be indicated by either a lighted sign or flashing red light to ensure that the appropriate safety glasses will be selected. The laser warning sign ideally is affixed to the room door using Velcro strips, which permit repeated attachment and removal without damage to painted surfaces. If the door is metallic, a magnetic pad may be glued to the back of the sign to allow repeated removal. Safety glasses should be available in a box outside the door for visitors not present at the beginning of the procedure.

The use of laser safety glasses must be mandatory when using wave guides, optical fibers, optical telescopes, and the microscope. Eye injuries from cracked fibers and malfunctioning mechanical shutter filters have been reported. Only when using CO_2 energy are the lenses within the optical telescopes and the microscope sufficient to prevent corneal injury.

The best protection against inadvertent laser hits away from the target area is coordination of laser actuation and standby modes. Any pause in laser activity should be followed immediately by a depression of the standby switch. The surgeon always must have an assistant to operate the laser mode and have the standby mode actuated as soon as active lasing is suspended. For example, when working in the oral cavity or nasal cavity, the laser should be in the standby mode as soon as the foot pedal is no longer depressed and prior to removal of the laser handpiece or fiber from the treatment area. Furthermore, when the laser is in the standby mode, confusing the laser foot switch with the foot pedal for the electrocautery or bipolar cautery will not cause any harm.

Eye and Skin Protection

Eye injury either to the cornea (focal cataract) or retina (blind spot or permanent blindness) can occur from a direct hit from the laser beam, a side hit, or a reflected beam hit. The specific injury will depend entirely on the wavelength of the laser energy and the energy density of the beam. Laser energy in the visible (argon, KTP, HeNe, etc.) and near infrared (Nd:YAG) regions of the electromagnetic spectrum (400–1400 nm) will tend to pass through the cornea and cause a retinal burn, whereas the CO_2 laser in the far infrared spectrum will be absorbed by the water content of the cornea and cause a thermal injury (cataract/scar). Protective eyewear dedicated to the specific wavelength of the energy being used must be available for the patient, the surgeon, and any other personnel present in the operating or treatment room. The safety glasses should have side shields to prevent a side hit and be ANSI-approved, with their optical density and wavelength stamped on the frame. Although most plastic lenses are acceptable for the CO_2 laser, glass lenses tend to resist scratching and will better withstand a laser hit. Laser safety glasses are essential for eye protection, but deeply colored lenses used for visible and near infrared lasers can obscure the operative field either by inadequate light or by color distortion. The operating surgeon must correct for these visual impairments and resist the temptation to operate the laser without safety glasses. Any of the currently available visible or near infrared lasers are capable of emitting a laser beam of sufficient power and within a time frame of such short duration that an aversion reflex or blink will not be sufficient to prevent an irreversible retinal injury.

Protection of the patient's eyes from the laser is also mandatory. Besides wearing appropriate laser safety glasses, patients should keep their eyes closed during laser use and be additionally protected with moist saline-saturated eye pads whenever possible.

Facial skin burns represent the second most frequently reported complication. This complication occurs when the skin is inadequately protected and the laser beam is misdirected, either by user error or specular reflection from a metallic instrument in the surgical field. The use of a double layer of saline-soaked gauze sponges, lap pads, or towels is effective for the patient under general anesthesia. However, for office-based laser surgery on an awake patient, this is seldom used. It is therefore crucial that the patient hold perfectly still during the procedure. For example, a rapid head turn during intraoral laser-assisted surgery could result in a facial or lip burn. Careful patient education with a videotaped demonstration of the procedure is quite helpful and reinforces the importance of a team approach.

Laser Plume Biohazard and the Need for Universal Precautions

Although stringent regulations concerning eye protection have been established by various regulatory agencies, there has been little written about protection from smoke inhalation during laser use. There has been increasing concern within the medical community that the noxious smoke plume generated by laser surgery may represent a significant health hazard both to the patient and to operating room personnel. The composition of plume generated during use of the CO_2 laser has been shown to contain particles ranging in size from 0.10 to 0.80 μ. These particles are too small to be effectively filtered by currently available surgical masks, although there are some commercially available laser masks that claim higher efficiency in filtering. Independent studies have shown some of these masks to be no better than standard surgical masks. Several studies have shown the laser plume to contain numerous gases and hydrocarbons that are toxic, potentially mutagenic, and carcinogenic. Bacteria and viral particles also have been isolated and may remain viable for up to 72 hours. In addition to obscuring the operative field and causing reflection of laser energy, laser smoke can be irritating to the eyes. Inhaled particles from laser plume fall into the size range of "lung-damaging dust" or particles that can travel to the most peripheral parts of the lung parenchyma. Inhalation of laser-generated smoke may cause transient nausea, hypoxia, depression of pulmonary defense mechanisms, and delayed airway inflammation. It is conceivable that repeated long-duration exposures to unfiltered laser plumes cause or exacerbate existing lung disease or cause the commonly termed black lung disease. Live viral deoxyribonucleic acid (DNA), including papilloma viruses, hepatitis viruses, and human immunodeficiency virus (HIV) have been isolated from the airborne contents of the laser plume and may be infectious. Therefore, laser plume must be treated like any body fluid that may contain bloodborne pathogens, and universal precautions should be instituted. This should be considered when handling and disposing used smoke evacuator filters.

Smoke Evacuation

The best way to minimize the risk of smoke inhalation during laser surgery is to ensure adequate smoke evacuation at the site of generation. This will prevent the generation of airborne particulate matter and eliminate direct inhalation by the surgeon, patient, or operating room assistant. Specially designed laser handpieces, nasal specula, laser fiber holders, and metal tongue depressors must all incorporate smoke evacuation channels. When not using one of these instruments, the smoke evacuator tubing should be connected to a pool tip suction (to prevent suction of skin or sponges) and held as close to the operative site as possible without interfering with the laser beam. The instruments are connected via tubing to high efficiency particulate air (HEPA) laser smoke evacuators capable of eliminating 99% of all particles in the range of 0.1 μ. To maintain high efficiency and prevent malfunction, filters must be changed according to the manufacturer's recommendations, and blood or secretions must never be suctioned into the smoke evacuator tubing. When working in the upper airway on an awake patient, laser treatment should be coordinated with breathing, so that tissue vaporization takes place while the patient exhales. This will divert the smoke plume toward the evacuation aperture of the laser instrument and prevent smoke inhalation.

Adequate room ventilation is essential and must exchange and filter the room air according to local standards. The laser generates heat, and in a small treatment room, the addition of fans and freestanding HEPA air cleaners may be needed. Ventilation ducts and vents should be positioned strategically relative to the treatment chair or table to maximize smoke evacuation.

SPECIFIC LASER APPLICATIONS IN THE HEAD AND NECK

Cutaneous Applications

Skin laser treatment may be performed in an office setting using topical anesthesia and several highly specialized new laser technologies. In an attempt to improve initial treatment results for vascular skin lesions (e.g., telangiectasias, port wine stains, hemangiomas, and venous lakes), pulsed lasers with more selective chromophore or pigment absorption spectra were developed to replace the KTP, argon, and Nd:YAG lasers, which often caused scarring or hypopigmentation. The "yellow" wavelength (577–585 nm) dermatological lasers (copper vapor and flash lamp pumped-dye lasers) evolved by necessity for treatment of vascular skin lesions. These pulsed lasers allowed a higher peak instantaneous delivery of energy than the average power. By pulsing the energy, decreased thermal spread and tissue injury are possible by using the heat sink effect of the circulation between pulses.

A more recently introduced generation of lasers is now capable of delivering high-intensity, short energy bursts that can pass through the epidermis with minimal superficial injury and permit even more selective wavelength absorption of either melanin (694 nm) or oxyhemoglobin (585 nm). The Q-switched YAG, Q-switched ruby, or Q-switched alexandrite (694) lasers are more melanin pigment selective, and the candela flash lamp-excited-dye (585) laser is more hemoglobin selective. These newer pigment-selective lasers have significantly improved the treatment of a variety of superficial skin lesions by reducing scarring and hypopigmentation, problems occasionally encountered in the past with argon and KTP lasers. Although many pigmented skin lesions can be cured or lightened with the laser, suspicious or recurrent lesions are still best treated by careful excision.

The use of lasers to achieve reproducible and predictable skin resurfacing has become a major component of cosmetic surgery and is performed routinely by dermatologists, general plastic surgeons, and facial plastic surgeons. The development of computerized, pulsed or scanning technology, for example, Silktouch (Sharplan Lasers, Inc., Allendale, NJ) and Ultrapulse (Coherent Inc., Palo Alto, CA), has added a new dimension to laser surface ablation. By pulsing the laser beam or scanning the laser beam over a defined area of skin, the laser's energy can be delivered at a relatively high power for brief periods of time. This results in superficial surface vaporization without char or thermal injury. Control over surface tissue ablation can now be achieved with a degree of precision not previously possible. Although chemical exfoliation, or chemical peel, continues to be a popular method for skin rejuvenation, more control over the desired depth of resurfacing is now possible when the CO_2 laser is coupled to this new technology. The CO_2 laser can be used to exfoliate the epidermis and dermis by superficial vaporization to renew sun-damaged skin and to ablate superficial, static wrinkles. Deeper, dynamic wrinkles or skinfolds formed by the underlying muscles of facial expression cannot be treated effectively by the laser but are amenable to botulinum toxin injections, selective muscle lysis, and skin tightening procedures (e.g., rhytidectomy and forehead lifts). The CO_2 laser is being used to debulk large keloids and mild to moderate acne scars and to "resurface" other skin areas such as superficial traumatic scars, pigmented solar keratoses, xanthelasma, spider hemangiomas, and other vascular and benign tumors. Laser treatment of rhinophyma, a particularly deforming sebaceous skin lesion of the nose, has also met with much success.

Laryngeal and Tracheobronchial Applications

The use of the CO_2 laser in otolaryngology procedures was first introduced in the early 1960s by Jako and Strong. The laser provided a unique tool for bloodless, sterile, no-touch microlaryngeal surgery for a variety of lesions. The management of recurrent laryngeal papillomatosis, webs, scar and stenosis, and telangiectasias are often best managed by combining traditional microsurgical techniques with the laser. The popularity of the laser for benign vocal cord lesions has declined, primarily because of accruing evidence of vocalis muscle scar and vocal cord stiffness. However, newer microspot focusing lens systems and application of microflap, mucosal welding techniques at low power settings (milliwatt) have refined use of the laser in phonosurgery.

Although there is continued controversy concerning postoperative voice quality, the use of the laser for ablation of early vocal cord cancers is a welcome and oncologically sound addition to open conservation laryngeal procedures and primary radiation therapy. In many situations, the transoral laser excision of superficial vocal cord cancers can be curative, with the lowest morbidity of any treatment modality. The laser is often useful for debulking obstructing laryngeal and endobronchial tumors. The CO_2 laser is usually used for the larynx and subglottic airway and the Nd:YAG laser for bronchial lesions. The Nd:YAG laser can be used with a fiber passed through a flexible bronchoscope.

Oral Cavity and Oropharyngeal Applications

Tumors of the tongue, gingiva, floor of the mouth, palate, and tonsils can be resected with excellent hemostasis and control using the CO_2 laser. Lack of muscle stimulation from the electrocautery can be of benefit in controlling tongue tumor margins. Because large vessels in the oral cavity cannot be controlled with the CO_2 laser, the electrocautery must always be on the field. For palliation of bulky, nonresectable tumor, the Nd:YAG can accelerate tumor necrosis and help debulk obstructing lesions or control bleeding from raw tumor beds.

The use of the CO_2 laser for tonsil surface ablation can be very effective for controlling or eliminating food collection in deep tonsillar crypts. It can be performed in the office with local anesthesia. Laser-assisted uvulopalatoplasty has become a popular office-based procedure for the management of habitual snoring and mild cases of obstructive sleep apnea.

Figure 14–9 *An intraoperative endoscopic photograph of laser turbinectomy. In this patient, an obstructing concha bullosa was vaporized with the Nd:YAG laser passed through a laser fiber, as seen in the foreground.*

Laser-Assisted Intranasal and Paranasal Sinus Surgery

Laser inferior turbinate reduction is most useful for submucosal intraturbinate introduction of a laser fiber to cause scarring and shrinkage of obstructing inferior turbinates. The Nd:YAG, CO_2, KTP, and diode lasers have all been used with good short-term and variable long-term relief of nasal airway obstruction (**Fig. 14–9**). There is considerable immediate postoperative edema and delayed tissue necrosis, often necessitating endoscopic debridement over a period of several weeks. The vaporization of small nasal polyps, vascular neoplasms, and synechea is easily achieved with the CO_2 laser passed through a wave guide under endoscopic guidance. Occasionally, obstructing cartilaginous septal spurs may be vaporized with the laser. The Nd:YAG laser has been an effective treatment for recurrent epistaxis, especially when caused by telangiectasias or other vascular lesions such as pyogenic granulomas.

Laser-assisted sinus surgery is limited by the inability of most lasers to ablate bone. The holmium:YAG (Ho:YAG) laser holds some promise for use in sinus surgery because it can be transmitted through a flexible fiber and will ablate bone. However, its safety near the orbital contents and brain has not yet been adequately studied, and long-term scarring may be a limiting factor. For the present time, most laser application in sinus surgery is limited to turbinate ablation.

Ear Applications

The argon, KTP, and CO_2 lasers have been very effective in stapedotomy, ossicular fixation, tympanosclerosis, and lysis of scar. The application of laser energy to the middle ear can be delivered via the microscope or through handheld fibers, depending on the wavelength used. The ultimate utility of the laser in middle ear surgery is still being evaluated but holds promise for the future. Minimally invasive flexible fiberoptic telescopes passed through a myringotomy eventually may be combined with miniature laser fibers for middle ear surgery without tympanotomy.

Laser myringotomy is a rapid and nearly painless procedure, useful for the management of otitis media with effusion in adults with poor response to medical management. A predefined puncture size can be created and will determine the length of time needed for complete healing.

Photodynamic Therapy

Light phototherapy is not a new concept. Sun exposure was used during early Egyptian times for a variety of skin ailments. Its contemporary use in the treatment of neonatal jaundice is well established. After systemic photosensitizing agents (hematoporphyrin derivatives) were introduced as a method to treat superficial tumors, the laser became an ideal light source for local activation. When transmitted through fiberoptics or endoscopes, laser light can reach remote intraluminal epithelial surfaces, delivering intense, monochromatic light for selective absorption by superficial tumors. Many pharmaceutical dyes and several intrinsic cellular constituents absorb light and undergo photochemical reactions, resulting in tissue destruction. The mechanism of tumor kill has yet to be fully elucidated, but it is most likely multifactorial, including direct oxidative alterations in enzymatic activity, generation of singlet oxygen-free radicals, and disruption of vascular endothelium causing thrombosis of the microcirculation. Because of their nonselective nature, all intravenously administered photosensitizing agents result in systemic toxicity mainly in the form of skin photosensitivity for variable periods of time postinjection. This necessitates thorough skin and eye shielding from all natural and artificial light for more than 6 weeks. This cumbersome and life-restricting precaution has been ameliorated somewhat by the development of newer laser dyes that are more selectively absorbed and retained by the target tumor tissue, and more easily metabolized and/or rapidly washed out of the nontumor tissue.

Photodynamic therapy has been most successful in treating superficial epithelial tumors of the genitourinary and gastrointestinal tracts, as well as skin and head and neck tumors. More specifically, early-stage tumors of the larynx and oral cavity have been cured with this technique. Because the tissue effect involves nonionizing radiation, it can be administered repeatedly for maximum benefit. Optimal laser wavelength and dosimetry continue to require further study, and not all tumors are equally responsive. Photodynamic therapy also has been useful as adjunctive multimodality management of head and neck tumors as well as for palliative treatment of more advanced, unresectable disease. The treatment of laryngeal papillomatosis with photodynamic therapy has been very exciting, with a few reported cures. Unlike the laser that causes submucosal scar and webs, the application of photodynamic therapy permits conservation of normal tissue and repeated application, making it an ideal treatment modality for this disease. Superficially located tumors in the upper aerodigestive tract can be treated effectively after uptake and retention of photosensitizing agents. Laser illumination can be repeated for maximal effect on tumor necrosis, with minimal local tissue inflammatory effects. The main limitations of photodynamic therapy include the need for total abstinence from ambient light exposure for prolonged periods after treatment due to skin photosensitivity, limited light penetration for deeply invasive tumors, and better, less toxic photosensitizers.

SUMMARY

Initially, the laser was a device in search of a disease to treat. However, novel applications, technological advancements, and new permutations of laser energy have established it as an important and irreplaceable surgical tool. There is no doubt that its role in minimally invasive endoscopic surgery and in future novel applications such as tissue welding and skin resurfacing will further solidify its position in our surgical armamentarium. The volume of ambulatory and, more specifically, office-based laser surgery will continue to escalate as pressure from managed care forces physicians and patients to accept its role in lowering the cost of health care in this country. It is our responsibility as physicians to ensure the safety of both our patients and our coworkers. Without the protective surveillance formerly provided by the hospital laser safety committee, it is now more critical than ever that laser safety guidelines be implemented and monitored in our medical offices and outpatient facilities. Furthermore, because laser-assisted surgery exposes the patient to a small but additional risk of injury compared with traditional techniques, its use must have clear advantages and must always be in the best interest of the patient.

Suggested Readings

Jako GJ, Strong MS. Laser surgery in the larynx: early clinical experience with continuous CO_2 laser. Ann Otol Rhinol Laryngol. 1972;81:791–798

Kuriloff DB. Laser safety in office-based ambulatory surgery. In: Yosef P, Krespi, eds. Office-Based Surgery of the Head and Neck. New York: Lippincott-Raven; 1998:3–13

Parkin JL. Lasers in otolaryngology. Otolaryngol Clin NA 1990; 23:1–169

Shapiro J, Zeitels SM, Fried MP. Laser surgery for larynx cancer. Operative Tech Otolaryngol Head Neck Surg 1992;3:84–92

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. Choose as many correct answers as apply to the question. To check your answers, see Answers to Self-Tests on page 716.

1. What are the most important unique features of laser energy that make it different from ordinary light?
   - A. It is coherent.
   - B. It is a form of ionizing radiation.
   - C. It is monochromatic.
   - D. It is polychromatic.
   - E. It is uncollimated.
   - F. It is collimated.
2. Which of the following are true about the differences between the carbon dioxide laser (CO_2) and the neodymium:yttrium-aluminum-garnet (Nd:YAG) laser?
   - A. The CO_2 laser is most useful as a surgical scalpel because of its precision and the fact that it is primarily absorbed by the surface water content of the tissue.
   - B. The Nd:YAG laser is most useful as a hemostatic laser because of its depth of penetration and ability to coagulate larger blood vessels than the CO_2 laser.
   - C. The CO_2 laser is more likely to damage the retina rather than the cornea of the eye if directly struck by the beam.
   - D. The Nd:YAG laser is an easier laser to learn with because of its hemostatic properties.
   - E. Unlike the CO_2 laser, the Nd:YAG laser requires a second laser just for aiming the treatment beam.
   - F. The safety glasses for the Nd:YAG and the CO_2 laser can safely be used interchangeably.
3. Which of the following statements are true about laser tissue interaction?
   - A. The argon laser cannot be used to treat bladder lesions because the energy will not pass through water without significant attenuation.

B. A highly focused spot size, short duration pulses, and very high power setting are the ideal laser parameters one should use to avoid excessive lateral thermal tissue damage and char.

C. The ideal laser parameters for surface tissue coagulation include a defocused beam, use of a lower power setting, and a laser absorbed by hemoglobin.

D. The power density of the laser can be changed by altering the spot size, focal point, and transmission through an optical fiber.

E. Flash scan technology was developed to increase the depth of penetration of the CO_2 laser to treat pigmented skin lesions.

F. The laser has replaced the surgical scalpel because of its ease of use and low cost.

4. Which of the following are important safety measures for using the laser?

A. The skin is always protected by the use of soaking wet towels.

B. Eye protection is needed for the Nd:YAG and CO_2 lasers because the laser beam is invisible.

C. Pure oxygen is used instead of nitrous oxide to decrease the risk of combustion from a stray laser hit through the endotracheal tube.

D. Eye protection should include glasses or goggles with side shields with the specific laser wavelength inscribed.

E. Smoke evacuation is not needed because viral particles vaporized into the laser plume are sterilized and can never infect personnel near the operative field.

F. Instruments used near the path of a laser should be ebonized or of low reflectance to reduce the risk of reflected laser energy striking a nontreatment target.

5. Certification for the use of the laser in medicine and the establishment of a laser credentialing committee is important for which of the following reasons?

A. The laser is a relatively new tool in surgery and medicine, with unique patient and operator hazards.

B. Because the laser is a new procedure, a committee must be established to set a realistic fee schedule.

C. The laser is so popular that its use must be triaged only for life-threatening procedures as determined by the laser committee.

D. Because the laser is expensive and often not covered by managed care insurance plans, the laser committee must defend its use for every procedure.

E. Because tissue interaction is unique for each laser wavelength, its use as a surgical tool and safety issues require an in-depth period of training through both didactic and hands-on training in the laboratory.

F. To ensure the proper and safe functioning of the available lasers in the operating room, their maintenance and review of safety procedures must be routine and performed on a regular basis.

Chapter 15

MOLECULAR BIOLOGY FOR THE OTOLARYNGOLOGIST

JEFFREY WOLFE, HINRICH STAECKER, AND THOMAS R. VAN DE WATER

The discovery of the double helical structure of deoxyribonucleic acid (DNA) by Watson and Crick laid the foundation for the field of molecular biology. With the advent of reverse transcription, the competitive reverse transcription (cRT) with ribonucleic acid (RNA) polymerase, polymerase chain reaction (PCR) amplification of DNA, and the complete mapping of the human genome, the scientific and medical communities have begun to establish a molecular basis for human diseases. Even more exciting is the theoretical ability to diagnose, treat, and possibly cure diseases on a molecular level. As the technology of molecular biology advances, it will likely begin to play a more prominent role in all fields of medicine, including otolaryngology–head and neck surgery. It is therefore crucial for the otolaryngologist to understand the basis of molecular biology and molecular genetics (see Chapter 19). The objective of this chapter is to introduce the reader to basic concepts in molecular biology and molecular genetics and to provide insight into the techniques used in these rapidly developing fields of science.

GENE EXPRESSION

The human somatic cell contains 46 chromosomes, of which half are inherited maternally and the other half paternally. Although all somatic cells in the body contain the same DNA structure, not all genes are expressed in every cell. As cells become differentiated in both structure and function, many genes that play important roles during development are turned off, leaving only those genes that allow for cellular preservation and function expressed. These differences in the pattern of a cell's gene expression result in many different cellular phenotypes that together make up the human body.

If human DNA were to exist as a single continuous strand, it would have an aggregate length of ~1.8 m. However, proteins called histones allow for compaction of this genetic material so that it fits within the nucleus of a cell. As cells undergo mitosis, the chromosomes unravel and lengthen. For cell division to occur, cells must duplicate their genetic material as well as their intracellular organelles. The cell cycle is divided into five phases (**Fig. 15–1**). The first of these phases is called gap 1

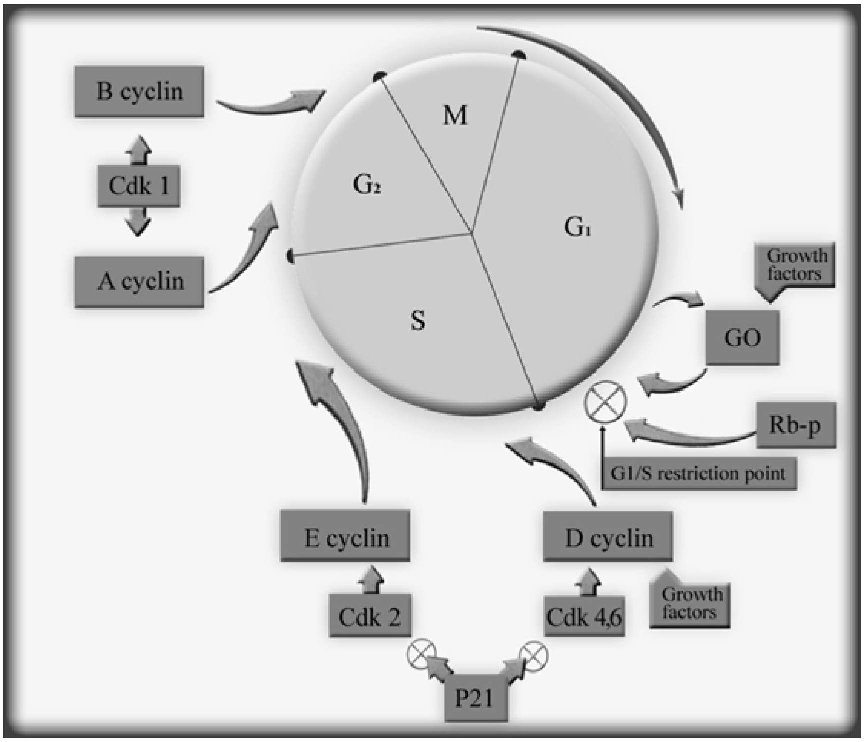


Figure 15–1 *Cell cycle basics. Control of the cell cycle is integral to cell growth and genesis of malignancies.*

(G1), and during G1, nucleic acids, enzymes, and other factors that guide cell division are produced. The synthesis phase (S) follows G1 and is the period of the cell cycle during which DNA replication occurs. After DNA replication is accomplished during the S phase, the gap 2 phase (G2) of cell division begins, and during this G2 phase, cellular proteins and intracellular organelles (e.g., mitochondria) are duplicated. The mitotic (M) phase follows the G2 phase and is when the replicated DNA is equally distributed as well as when the division of the parent cell occurs, forming two progenic daughter cells. Cells that are dividing rapidly may immediately revert to the G1 phase, whereas other cells that are entering a differentiation and maturation phase may leave the active cell cycle and enter a rest (G0) phase. Cells such as muscle cells and neurons spend most of their existence in the G0 phase, whereas rapidly dividing cells such as the epithelial tissue lining of the intestines spend very little time in G0 (**Fig. 15–1**). Malignant disease can result from a loss of internal control of the cell cycle. Negative regulatory genes that influence a cell's ability to reenter the cell cycle and code for proteins that suppress cell division are called tumor suppressor genes (anti-oncogenes).

The information in DNA is contained in a code of four nucleic acids: adenine (A), guanidine (G), cytosine (C), and thymidine (T). In the double helical structure of DNA, adenine binds to thymidine and guanidine to cytosine, which imparts structural stability to the helix. The production of proteins is accomplished by creation of messenger RNA (mRNA) via transcription and by translation of the encoded mRNA (**Fig. 15–2**). Transcription occurs via an enzyme called RNA polymerase II, and this process results in the production of a single strand of mRNA that is stabilized with a cap on the 5′ end and with a polyadenylated tail on the 3′ end. Transcription of a gene to be translated requires specific binding of the RNA polymerase II to a promoter, which can only occur if an appropriate translation factor has bound previously to a TATA box sequence. Each gene is composed of segments of DNA called exons and introns (**Fig. 15–2**). Exons are subunits that encode the information portion of the gene and are interrupted by intron sequences, which contain sequences that can alter the transcription of the gene in question and also function to allow for the creation of alternative configurations of a protein via the arrangement of different combinations of exons (i.e., splice variants). The mRNA created from the transcription of exons on the chromosomes leaves the nucleus and is translated within the cell's cytoplasm into a protein in the ribosomes via the combined act of both ribosomal (rRNA) and transfer RNAs (tRNAs).

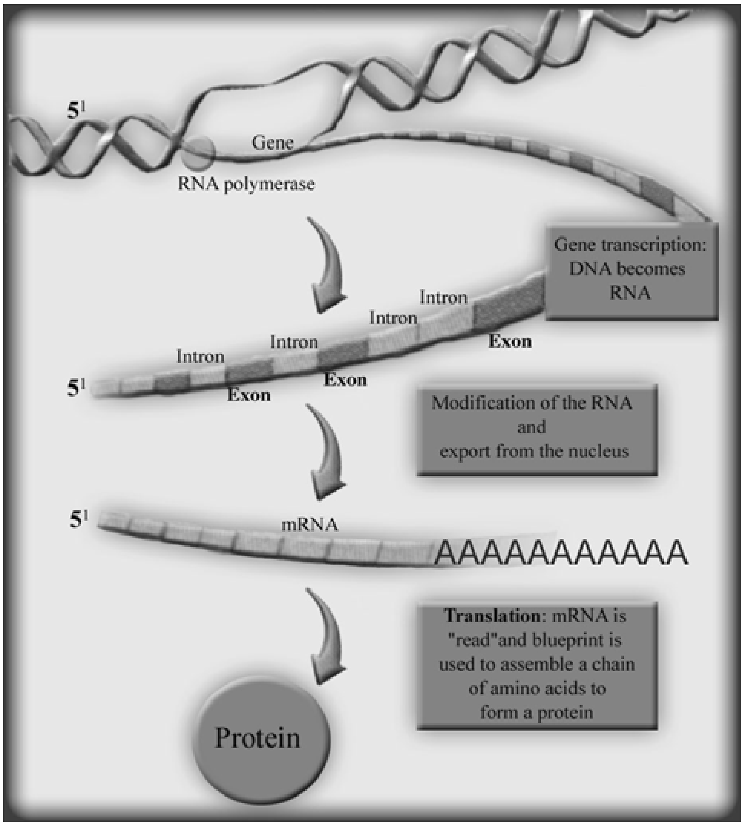


Figure 15–2 *Transcription/translation. Production of proteins depends on the copying of genes by RNA polymerase and the processing of the resulting messenger RNA within the cell.*

TRANSCRIPTION FACTORS AND REGULATION OF GENE EXPRESSION

Regulation of gene expression in both a time- and tissue-specific manner is a defining factor that specifies the phenotype of a cell. During differentiation, a series of genes that define a cells position and fate are active and then silenced as a cell matures and assumes an adult phenotype. Methylation and the positioning of nucleosomes that makes DNA inaccessible to the transcription machinery frequently accomplish this silencing. At the most basic level, a gene needs to be "read" by RNA polymerase II in order for gene transcription to occur. The efficiency and speed at which this takes place is controlled by upstream sequences of DNA called promoters. The TATA box is a sequence of DNA upstream to the 5′ initiation codon of a gene that has been found to be necessary for both the function and the binding of RNA polymerase. Additionally, TATA box binding factor is required for binding of the RNA polymerase to the initiation site of an exon sequence of DNA. Many different types of transcription factors have been identified and classified. Most of these factors have been evolutionarily conserved and are ubiquitous; however, some of these factors are fairly tissue specific. Among the most important of these transcription factors are the leucine zipper (fos/jun factors) and zinc finger proteins. Other DNA sequences known as enhancers can increase transcription of a given gene; these sequences have been found in both the 5′ and the 3′ regions of an exon and even within intron sequences that are located far away from the gene of interest. Growth factors can influence the rate of transcription by indirectly increasing the action of either or both transcription activators and transcription inhibitors. This is commonly accomplished via a cascade of phosphorylation of a class of signal proteins called kinases that act upon either tyrosine or serine/threonine sites within targeted molecules.

ONCOGENES

Oncogenes and tumor suppressor genes (anti-oncogenes) alter cell growth in cancer. Increased expression of oncogenes and/or decreased expression of tumor suppressor genes can increase the rate of proliferation of tumor cells. Proto-oncogenes are genes that function in the cell growth and tissue repair of normal tissues, but when they are abnormally activated via mutation, they can stimulate uncontrolled cell proliferation. *P53* is a tumor suppressor gene in which mutations are often found in many types of cancer, including those of the colon and of the head and neck. Mutations of the *p53* gene are found in over 40% of head and neck squamous cell carcinomas.

TECHNOLOGY IN MOLECULAR BIOLOGY

Fundamentals of Gene Manipulation and Isolation

The overall human genome is estimated to contain ~30,000 to 40,000 protein-coding genes within 3.5×10^9 bp of DNA nucleotides. Techniques to isolate and study these genes are all based on techniques developed from studies with bacteria. DNA can be cut into segments by a variety of enzymes known as restriction endonucleases. These enzymes recognize a specific sequence of four to eight nucleotides and cut the DNA at this site, leaving either blunt or staggered ends. The cut DNA can then be pasted (cloned) into a small circular bacterial DNA called a plasmid. Plasmids carrying a cloned foreign DNA of interest and a gene for antibiotic resistance (selection gene) are then transferred into bacteria that are grown on agar containing an antibiotic, thus allowing for selection of bacterial colonies carrying the foreign DNA. As an adjunct to this gene detection procedure, several different gene detection techniques are available that take advantage of DNA's propensity to form a helix with a complementary molecule. If the sequence or part of a sequence of a protein is known, a complementary strand of nucleic acid can be synthesized and labeled radioactively. DNA from bacterial colonies or a DNA size sequence ladder from an agarose gel can be transferred onto a nitrocellulose support membrane. Heating of the DNA causes the hydrogen bonds of the double strands to open. Cooling them in the presence of the probe will cause the formation of a new helix consisting of a native DNA strand and the strand of radioactive tagged probe DNA. These can then be detected by autoradiography. This process is known as hybridization and can be performed on a nitrocellulose filter membrane or on a tissue section to yield localization of a gene to a tissue sample or within a tissue, respectively. Alternate strategies of cloning involve isolating all the mRNA from a particular cell type. DNA copies of the mRNA are then made using an enzyme, reverse transcriptase, yielding a complementary (c)DNA library that can then be probed for a particular sequence of DNA that represents an expressed gene.

There is limitless potential from DNA technology. Aside from isolating and synthesizing a large number of copies of a gene, one can design plasmids that allow for the expression of an inserted gene and thus initiate the blueprints for production of a specific protein; cDNA that encodes a gene can be cloned into a plasmid that carries a promoter sequence that is effective for either bacterial or eukaryotic cells. Insertion of a plasmid into bacteria results in the production of a large amount of recombinant protein. Alternately, insertion of the plasmid into a eukaryotic cell line allows one to study the change in a cell's phenotype induced by overproduction of a normally occurring protein or the expression of an inappropriate protein within a specific type cell. Thus one could introduce and overproduce a gene into a cell line to determine its function. Recent advantages in recombinant gene technology have resulted in the development of several viral vectors. Viral vectors allow for more efficient transfer of DNA than by the insertion of plasmids, and these vectors have powerful promoter systems that can be selected to target specific classes of cells (i.e., target a neural cell by using a neuron-specific enolase promoter to drive the transcription of the vector-inserted gene). The viral genes responsible for the lytic cycle are removed and replaced by the specific gene that is to be transferred. The modified viral genome is then packaged in a cell line that produces the viral capsid proteins. The viral particles can be isolated, concentrated, and used to transfer and express a gene of interest in a host tissue. Currently, most viral vectors are based on herpesvirus; adenovirus, adeno-associated virus, lentivirus, and other retroviruses (RNA viruses). If the gene therapy vector remains as an episome and does not integrate into the host DNA, the process is termed transduction (e.g., adenoviral and herpesviral vectors). However, if the vector integrates into the host cell's DNA, the process is called transfection (e.g., adeno-associated viral, lentiviral, and retroviral vectors). As a general principle, transduced genes are expressed for a shorter period of time, and transfected genes are expressed for a longer period.

Polymerase Chain Reaction

The discovery and development of the PCR has significantly accelerated progress in molecular biology. This technique depends on the action of Taq polymerase, which is heat stable and remains active at the high temperatures (i.e., 70–75°C) required to extend a single DNA strand after its denaturation and separation from double-stranded DNA is accomplished at 94 to 95°C. Using two short oligonucleotide primer sequences from the 5′ and the 3′ regions of the gene of interest, Taq polymerase is able to synthesize the remaining complementary strand of DNA after the primer has formed a complex with a single strand of native DNA. The reaction mixture of DNA, oligonucleotide primers, and Taq polymerase is then reheated in a programmable thermocycler to repeat the cycle, resulting in exponential amplification of the sequence that lies between the two flanking oligonucleotides. Ultimately, the sequence of the DNA can be determined from the amplified DNA, yielding an exact comparison to normal wild-type DNA. This technique can be used for cloning or for testing for the presence of and fidelity of any specified sequence of DNA within a tissue. Detailed information regarding the presence of mutations can then be obtained from automated DNA sequencing or a variety of techniques that use the melting temperature characteristics of DNA to determine homology between a control and an amplified sample strand of DNA. The PCR can be used in several different ways beyond the screening of cellular DNA. The ability to use PCR technology also exists for studying expressed genes by assaying for the RNA message (mRNA) using reverse transcription PCR (RT-PCR). After isolating total cellular RNA from a tissue sample, reverse transcriptase can be used to synthesize DNA copies of any mRNA present. PCR amplification of these DNA copies can then be performed to determine if a gene is expressed in certain cells and tissues. This technique has also been applied to tissue sections, allowing detection of very low copy numbers of mRNAs that were under the limit of normal detection within specific sensitivity of tissues in situ hybridization. The technique can also be used to determine relative levels of gene expression.

Site-Directed Mutagenesis

The effect of nucleotide sequence on gene and protein function can be determined by inducing specific mutations into a gene using site-directed mutagenesis, which relies on specifically designed oligonucleotides that complement a known sequence. An oligonucleotide carrying a specific mutation (point mutation) is annealed to wild-type single-stranded DNA. Viral or bacterial polymerase is then used to make copies of selected DNA, creating a population of normal and mutated DNA. This can then be transfected into a cell line and the cells screened for the effect of the mutation on their phenotype and physiological function.

On a much larger scale, the use of classical genetics and cytogenetics has helped us to understand the importance of the inheritance of cancer susceptibility and to study chromosomal damage directly. Localization of a gene can be examined either through linkage, which statistically establishes the distance between a genetic trait and known chromosomal makers, or one specific chromosome through physical mapping of the position of the gene. Restriction fragment length polymorphisms (RFLPs) have been used to generate a restriction map that shows the arrangement of a specific set of restriction sites within a segment of DNA and therefore can be used to follow or track the inheritance of a part of a chromosome through a family. After generating a family tree that describes phenotypes, DNA is obtained from each of the family members and then is digested with a restriction enzyme and hybridized with a probe to a template-specific gene sequence or marker via Southern blotting that is known to correlate with the disease. Differences in the pattern of restriction digestion of the marker gene or sequence of DNA from a known chromosomal location are then statistically correlated to the presence or absence of the disease. If two genetic sequences are on different chromosomes, their inheritance will be unlinked; that is, they have only a 50% chance of being inherited together. The closer together two genes are on a chromosome, the greater the probability that they will not be separated by crossover events and will therefore be co-inherited. By screening large family groups for the co-inheritance of a certain gene and a large number of individual RFLPs, a few RFLP markers can be identified that surround the gene and DNA sequences in the region of the gene can be located, and the DNA that corresponds to the gene itself can be located. Currently, mapping and cloning studies are increasingly relying on the use of single nucleotide polymorphisms (SNPs) as tags or linkage markers for genes of interest. SNPs are highly abundant polymorphisms that occur approximately every 1 in every 1000 base pairs (bp) of DNA, and therefore can be used for both linkage and association mapping of inherited characteristics. Another method to detect mutations is via karyotyping. Chromosomes are grouped by size, centromere position, and banding pattern when stained with dyes specific for DNA. By convention, the short arm of a chromosome is labeled *p* (petite), and the long arm of the chromosome is labeled *q*. Using a known polymorphism in a linkage map, the frequencies of recombinations between genes and markers can also be related to the physical banding pattern of the chromosome. Physical mapping is an extension of karyotyping in which the banding pattern of normal metaphase chromosomes is compared with the location of the gene determined by hybridization of a probe to the whole chromosome. The combination of cytogenetics, which yields information regarding loss or translocation of a portion of the chromosome with physical mapping, has established many of the underlying molecular pathologies of neoplasms (e.g., *p53* mutations).

Gene Knockouts

One method of studying the function of genes is creation of mice in which the gene of interest has been disrupted (null mutation, gene knockout). The initial step involves the creation of a defective gene. The gene of interest is cloned into a vector that also carries the gene for herpesvirus thymidine kinase (TK). Next, the center section of the targeted gene is deleted, and a gene conferring antibiotic resistance is inserted in its place. This yields a segment of DNA that contains the gene for antibiotic resistance flanked by the DNA of gene X, followed by a stretch of unread DNA and then the TK gene. The recombinant DNA is then inserted into a culture of embryonic mouse stem (ES) cells in vitro. One of three outcomes can then happen: (1) homologous recombination occurs between the native gene X and the genetically altered gene X in the ES cells, (2) the recombinant DNA inserts randomly into the ES cell DNA, or (3) no insertion or recombination takes place. ES cultures are then treated with neomycin, killing all cells that did not take up the recombinant DNA or recombine with the modified gene X. Subsequently, the ES cells are treated with ganciclovir that selects against cells that integrated the gene for TK into their genome (i.e., only ES cells where random insertion of the entire DNA segment had taken place would die because ES cells that underwent homologous recombination would not carry the TK gene). The selected ES cells would carry one wild-type copy of the normal gene X and one copy of the mutated gene X interrupted by the inserted neomycin-resistance gene. These ES cells would then be reimplanted into early-stage mouse embryos, creating chimeras containing germ and somatic cells with one copy of each type of gene X. Cross-breeding of these genetically manipulated mice can then establish stable heterozygous breeding lines that can be crossed with each other to yield offspring that are homozygous (null/null) for the mutated gene X. The gene knockout animals (null mutants) will then portray the gene X −/− (homozygous) phenotype.

Tumor Bioassays and Cell Lines

The use of cell lines has been invaluable for the study of oncogenesis. Initially, cell lines were transformed by oncogenic retroviruses, allowing early analysis of the function and phenotypic changes in the cultured cells induced by the oncogenes. Normal characteristics that can then be reevaluated include growth characteristics, density-dependent growth inhibition, and serum and growth factor dependencies. Many well-characterized cell lines are available that allow the assessment of genes introduced into these for cell lines derived from a specific tissue. Foreign DNA can be transduced or transfected into a cell in several different ways. Plasmids carrying a gene of interest that is linked to a promoter can be introduced by either altering the medium's calcium concentration or treating the cells of interest with a weak electric field (electroportation). Alternately, a gene gun can be used to introduce foreign DNA into a cell by carrying the DNA on very small metal particles (e.g., gold) that are shot at high velocity into the cell. Finally, a series of viral vectors can be used to introduce and express (i.e., transfect or transduce) genes into mammalian cells.

Flow Cytometry

Many tumors have been shown to carry an abnormal complement of DNA, having tetraploid or polyploid genomes rather than diploid. Studies have observed that certain tumors with aneuploidy have worse prognoses than tumors that have a normal complement of chromosomes. The proportion of cells within a population that have an abnormal number of chromosomes (i.e., aneuploidy) can easily be determined using flow cytometry. In this technique, cells are stained with a fluorescent dye that colors DNA and then passed through a stream of laser light. Analysis of scattered light yields an estimate of cell size, whereas analysis of their fluorescence allows for the calculation of the total DNA content of that same cell.

Genomics

Automation of genomic analysis has accelerated the process of discovery. RNA extraction, PCR, and base pair sequencing can be performed by robots, allowing the rapid sequencing of the entire spectrum of mRNAs expressed in a given cell type or even sequencing the cDNA to produce genomic sequences. The sequencing of specific cellular cDNA libraries has been used to create databases of expressed sequence tags. Use of standard gene browser computer programs can lead to rapid identification of an unknown gene sequence by searching available databases.

The availability of known gene sequences in combination with robotic techniques has allowed researchers to develop "gene chips." In this technology, spots of unknown DNA are attached to a slide or membrane, and RNA isolated from a cell or tissue sample is hybridized to this membrane that can carry thousands of known sequences. The hybridization pattern seen reflects the expression pattern of genes within a cell or tissue. Thus normal versus cancer cells can be screened for differences in gene expression, leading to the formulation of potential molecular intervention strategies. A recent advance in molecular genetics is the serial analysis of gene expression (SAGE), which allows for the analysis of all transcripts from a group of cells or sample of tissue by creating a tagged 10 bp unique sequence for each transcript. Analysis of expression patterns can be performed by comparing sequences produced by SAGE analysis to established libraries, thereby allowing for the detection of even low copy genes that may not be picked up by standard gene chip analysis.

SUMMARY

The discipline of molecular biology is evolving rapidly, and applications to otolaryngology–head and neck surgery are keeping pace with this expansion. The basic biochemistry of DNA has allowed the development of sophisticated genomic techniques that can be applied to the study of the underlying mechanisms of many ear, nose, and throat disorders; for example, hearing loss and head and neck cancer. Future developments will focus on the application of bioinformatics to help us understand the vast amount of information that can be generated by these modern molecular techniques. Data on gene and protein function, in combination with studies of patients' diseases, will make molecular medicine a useful addition to the treatment of otolaryngological disorders.

Suggested Readings

Lodish H. Molecular Biology of the Cell. New York: WH Freeman; 2000

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Which of the following is not a technique used to analyze gene expression?
   - A. In situ hybridization
   - B. Polymerase chain reaction
   - C. Flow cytometry
   - D. Serial analysis of gene expression (SAGE)
2. An example of a gene identified as playing a role in the molecular generation of head and neck cancer is
   - A. Tumor necrosis factor α (TNF-α)
   - B. Tubulin
   - C. *p53*
   - D. DNA polymerase
3. Reverse transcription refers to
   - A. Amplification of genomic DNA
   - B. The use of an RNA polymerase to produce a copy of cDNA from transcribed messenger (mRNA)
   - C. Isolation of transfer RNA from a cell extract
   - D Localization of mRNA transcripts in a histological tissue sample

Chapter 16
PHYSIOLOGY OF THE PEDIATRIC PATIENT

Lewis P. Singer

Infants and children cannot just be considered small adults. There are differences in metabolism and in the maturation of almost all organ systems, from the cardiovascular and respiratory systems to the central nervous system. This chapter will describe many of these differences as they pertain to the pediatric airway, specifically of the premature infant, the term newborn, and young infants. Finally, physiological changes that accompany tracheotomy will be discussed along with concerns during decannulation.

During fetal life, respiratory effort is part of normal development, providing work for the diaphragm. Fetal pulmonary blood flow is minimal due to pulmonary arterial vasoconstriction. The lung bud appears by the end of the fourth week of gestation. Bronchi appear by the sixth week. Terminal bronchioles are completed by 16 weeks' gestation. Alveolar formation begins in the late third trimester. A full-term infant has 20 million to 50 million terminal air sacs, ~10% of a normal adult complement. Lung volumes in newborns are disproportionately small compared with body size. The ratio of metabolic rate to body weight in an infant is twice that of an adult, making ventilatory requirements higher for the infant with smaller lungs and giving infants much less respiratory reserve.

RESPIRATION

As lung growth continues for the first decade of life, most alveoli develop postnatally. With the first few breaths at birth, major changes take place in the lungs and the cardiovascular system. During the transition to extrauterine life, the liquid in the air spaces and airways of the lungs is rapidly absorbed and exchanged for air, leading to pulmonary vasodilatation and increased pulmonary blood flow. Cord clamping initiates continuous rhythmic breathing; however, normal neonatal respiratory adaptation and control will take several weeks. Continued and rhythmic breathing is maintained by normal arterial pH. Continuous breathing is independent of the partial pressure exerted by carbon dioxide dissolved in arterial plasma and red blood cell water ($PaCO_2$) and unaffected by carotid body denervation. Unlike hypoxia in older infants and children, newborn hypoxia does depress respiration and cause apnea.

The precise control of fetal breathing is uncertain. During the last 10 weeks of gestation, the fetus "breathes" ~30% of the time. The respiratory rate at 30 weeks' gestation is ~58 breaths per minute; at term the rate falls to ~47. There are acute increases in respiratory rate a few hours after a maternal meal, correlating with maternal blood sugar. Spontaneous fetal breathing occurs during rapid eye movement (REM) sleep and is independent of chemical and nonchemical stimuli. The normal fetal PaO_2 is 20 to 25 mm Hg, which normally would have an inhibitory effect on breathing and REM sleep. Severe fetal hypoxia ($PaO_2 < 10$ mm Hg), usually associated with fetal acidosis, results in gasping respirations. Fetal breathing may represent "practice" and is a stimulus to fetal lung growth. Experimental cessation of breathing causes hypoplastic lungs.

Respiratory function in the first months of life is very different from later childhood, adolescence, and adulthood. In young infants, the diaphragm is the most important muscle of respiration. These infants have an extremely compliant chest wall. The intercostal muscles stabilize the chest wall, counterbalancing the force of diaphragmatic contraction, so that the lungs can expand. Any disturbance that changes the relationship of these two forces increases the work of breathing for the infant. Muscle fatigue or weakness leads to chest wall collapse seen clinically as retractions of the chest wall during inhalation. Parenchymal lung diseases such as pneumonia and respiratory distress syndrome decrease the compliance of the lungs, and, in the face of a very compliant chest wall, the chest wall collapses, with a resulting major increase in the work of breathing. An infant, therefore, develops respiratory failure much earlier than an older child or adult, necessitating assisted ventilation. The infant lung is normally relatively noncompliant, tending to collapse as lung disease develops. Low lung compliance, along with a relatively compliant chest wall, results in a low functional residual capacity very close to residual volume. An infant compensates for this by narrowing the laryngeal airway during exhalation, an attempt to maintain positive intrathoracic airway pressure. As lung disease progresses, this becomes a more pronounced audible exhalatory "grunt." Grunting respirations are an important sign of impending respiratory failure.

Surfactant produced by type II alveolar cells is responsible for the decreased surface tension that is necessary to prevent alveolar collapse at small lung volumes. Inadequate or deficient surfactant is the etiology of respiratory distress syndrome (RDS) in newborns, common in premature infants and infants of diabetic mothers. A lack of surfactant leads to alveolar collapse, atelectasis, and decreased pulmonary compliance. Surfactant replacement therapy provides good clinical results and has been employed as treatment for both meconium aspiration syndrome and adult respiratory distress syndrome.

Control of breathing is maintained by the respiratory center in the brainstem, with input from central and peripheral chemoreceptors, and stretch receptors in the airways and lung parenchyma. In the brainstem the dorsal respiratory group of neurons is located along the tractus solitarius, which is the principal site for the ninth and tenth cranial nerves, which carry afferent nerves from the lungs, airways, heart, and peripheral arterial chemoreceptors. Ventilation is adjusted to maintain a normal blood pH between 7.35 and 7.45, while maintaining the arterial $PaCO_2$ in a very narrow range. This occurs no matter how oxygen consumption and carbon dioxide production change with metabolism. Inspiration is initiated by diaphragmatic contraction developing subatmospheric intrapleural pressure and resultant lung expansion. Exhalation is passive, resulting from elastic recoil of the lungs and chest wall.

Three afferent neurons have been described. Iα fibers are inhibited by lung inflation. Iβ fibers react to lung inflation and pulmonary stretch receptors. The third type, called P cells, receive impulses following lung inflation. Excitation of Iβ fibers is associated with shortening of inspiratory time and prolongation of exhalation time. This response is likely responsible for the Hering-Breuer reflex, which inhibits respiration during lung inflation. The ventral respiratory group is responsible for quieting inspiratory drive during exhalation, and the pontine respiratory group plays a secondary role in this response. Rhythmic breathing can occur in the absence of feedback from peripheral receptors. Transection of the pons has little effect on respiratory pattern, implying that the respiratory generator is likely in the medulla.

Peripheral receptors can be found in the upper airways, trachea, bronchi, lungs, and chest wall. Upper airway receptors are responsible for sneezing, apnea, changes in bronchomotor tone, and initiation of the "diving" reflex, which is the profound bradycardia associated with a shift of blood flow to the brain and heart thought to occur during submersion. Epipharyngeal reflexes produce sniffing, moving mucus from the nasal airway to the pharynx. Pharyngeal receptors facilitate swallowing, inhibit breathing, close the larynx, and coordinate contractions of the pharyngeal muscles. Laryngeal receptor activation causes apnea, coughing, and changes in the respiratory pattern.

Slow-adapting or stretch receptors are located in the submucosal muscles in the membranous portion of the trachea and are activated by distention of the airways. The Hering-Breuer reflex inhibits inspiratory activity, modulating respiratory effort. Hypocapnia augments these reflexes, and hypercapnia inhibits them. Apnea provoked by an inflated endotracheal tube cuff may be due to these slow-adapting reflexes.

Rapid-adapting receptors in the mucosa of the carina and large bronchi can be stimulated by irritant gases, anesthetic or toxic, and are responsible for initiating coughing, bronchoconstriction, and mucus production. C-fiber endings located in capillary walls are stimulated by pulmonary vascular congestion, pulmonary edema, microvascular emboli, and noxious gases, producing apnea and slow, shallow breathing, hypotension and bradycardia, bronchoconstriction and mucus production, and laryngospasm.

Newborn animals are very sensitive to stimulation of receptors innervated by the superior laryngeal nerve resulting in respiratory depression or apnea. This could explain the tendency of small infants to become apneic from any noxious substance in the larynx. Inhalation of anesthetic agents in small infants frequently produces coughing, breath holding, and laryngospasm.

Central receptors on the surface of the ventrolateral medulla are anatomically separate from the respiratory center. These receptors respond to changes in hydronium ion (H^+) concentration in the cerebrospinal fluid (CSF) and not to PCO_2 or pH. Carbon dioxide rapidly diffuses into the CSF, quickly determining H^+ concentration. Acute respiratory acidosis produces a rapid stimulation of these central receptors, whereas metabolic acidosis or alkalosis does not generate as rapid a response. During chronic respiratory acidosis, as seen in children with bronchopulmonary dysplasia, the CSF pH remains relatively normal. In this situation respiratory stimulation is more dependent on peripheral chemoreceptors that respond to hypoxemia. Oxygen therapy blocks this hypoxic drive, which puts these patients at risk for hypoventilation during oxygen administration.

Peripheral chemoreceptors are located in the carotid bodies and are responsible for the tachypnea that is present during hypotension. They also respond to changes in arterial pH and $PaCO_2$. An arterial PaO_2 less than 60 mm Hg results in a significant increase in ventilation, except in premature and term newborns, when hypoxemia results in respiratory depression. Anemia or carbon monoxide poisoning does not result in stimulation of the respiratory center because the arterial PaO_2 is usually normal under these conditions. General anesthesia, opioids, and sedatives depress ventilation, variably affecting tidal volume and respiratory rate.

Neonates and young infants develop a transient stimulation of respiration in response to hypoxia. They usually generate one good breath and then become apneic. The transient stimulation is abolished in a cool environment like an operating or delivery room. By 3 weeks of age, the response to hypoxia generally is just respiratory stimulation. Newborns respond to hypercapnia with hyperventilation, but with a somewhat attenuated response compared with older infants.

Periodic breathing and central apnea are respiratory patterns common in young infants. There is a concern that these breathing patterns associated with immaturity may be related to sudden infant death syndrome (SIDS), but this is as yet unproven. Periodic breathing is described as an irregular breathing pattern with periods of breathing interspersed with short periods of apnea from 5 to 10 seconds in length. This irregular breathing is usually associated with REM sleep. Full-term infants spend ~50% of sleep time in REM sleep as compared with 20% for adults. Premature infants remain in REM sleep most of the time and breathe irregularly. Periodic breathing may occur during wakefulness, active REM sleep, and quiet sleep. The incidence of periodic breathing is 78% in full-term infants and 93% in premature infants. Periodic breathing decreases with increasing age, with an incidence of 29% by 1 year of age. Breathing 2 to 4% carbon dioxide depresses periodic breathing by stimulating respiration.

Central apnea or apnea of infancy is defined as cessation of breathing for 15 seconds or longer, or shorter periods when associated with bradycardia, cyanosis, or pallor. This is more common in premature infants and is probably related to immaturity of the central nervous system. After simple surgical procedures such as inguinal herniorrhaphy, life-threatening apnea may occur, especially in premature infants who are less than 50 weeks postconceptual age. Theophylline and caffeine are effective central respiratory stimulants that are frequently used as therapeutic agents for apnea of infancy.

Intrapleural pressure is 5 cm below atmospheric pressure in older children and adolescents, compared with 0 or 1 cm below atmospheric pressure in newborns. The outward recoil of the infant chest wall is very low, resulting in a functional residual capacity near closing volume. Because end-expiratory lung volumes in infants are so near closing volume, conditions that promote airway closure and loss of lung volume will result in atelectasis and ventilation-perfusion mismatch. Apnea, general anesthesia, and muscle relaxation are examples of such conditions frequently encountered in the operating room and intensive care unit. Abdominal surgery, abdominal distension, and thoracic surgery have a similar effect. Newborns with RDS have very stiff, noncompliant lungs that are difficult to expand, reducing lung volumes further.

Resistance in the pulmonary system is composed of resistance of the lung to distension and resistance to airflow through the airways. During laminar flow, flow resistance can be described by Poiseuillle's law, where resistance (R) is proportional to the length of the tube (l) and viscosity of the gas (η), and inversely proportional to the radius (r). The units are expressed as pressure per flow or cm H_2O/L per second:

$$R = 8l\eta/\pi r^4$$

Clearly, radius of the airway has the largest effect on resistance to airflow. One millimeter of circumferential subglottic edema in an adult has a minor effect on airway resistance, whereas the same amount of edema in the small airway of an infant will cause a 16-fold increase in airway narrowing and resistance. Turbulence is much more common in the narrow airway of small infants. During turbulent flow, resistance is related to $1/r^5$. Breathing mixtures of helium and oxygen (at least 60% helium) decrease resistance in areas of turbulent airflow due to the lower density of helium compared with nitrogen. The nasal airway is the normal route of airflow into the respiratory system and comprises 55% of the total airway resistance. This is twice the resistance of one who mouth breathes. Placing a nasogastric tube will narrow the nasal airway even further and increase total airway resistance in an infant by 50%.

The pharyngeal airway is composed of soft tissues easily collapsed by negative intraluminal pressure, by posterior displacement of the jaw, muscular hypotonia, or paralysis. Mechanical support to maintain patency of the pharyngeal airway occurs by increasing muscle tone or contraction of the genioglossus, geniohyoid, sternohyoid, sternothyroid, and thyrohyoid muscles that make up the pharynx. These muscles synchrononously increase their tone along with diaphragmatic contraction during inspiration. Chemoreceptors responding to hypercapnia and hypoxemia also increase pharyngeal muscle tone.

The function of the larynx, the organ of phonation, is to maintain the airway open and to occlude and protect the lower airway from substances during swallowing. The cricoid cartilage forms a complete ring, preventing the proximal tracheal airway from complete collapse. The infant's larynx is high in the neck at C3–C4 and is funnel shaped, with the smallest part inferiorly at the cricoid cartilage. The epiglottis is Ω shaped and juts out at a 45-degree angle. With growth, the larynx descends to C5–C6, becoming cylindrical in shape, with the cords forming the narrowest portion. In adults, the epiglottis is thin, arises close to the base of the tongue, and generally parallels the course of the trachea. The cricoid cartilage surrounds the subglottic space that is easily traumatized from oversized or tight endotracheal tubes. Tight-fitting endotracheal tubes may cause mucosal ulceration and pressure necrosis, and inflammation of the cricoid cartilage. The ischemic injury may result in mucosal edema that is responsible for postintubation croup. Subsequent fibrosis of the cricoid cartilage may lead to subglottic stenosis.

Normally, the glottis widens during inspiration, minimizing airflow resistance, but narrows during expiration that maintains positive airway pressure at end expiration and residual lung volume, the functional residual capacity (FRC). This function is lost with endotracheal intubation or tracheotomy due to loss of glottis closure. Application of positive end-expiratory pressure (PEEP) will help maintain FRC.

There are many protective reflexes. Coughing, sneezing, swallowing, and laryngeal and pharyngeal closure are all attempts to maintain the airway open and clear. Laryngospasm is marked closure of the glottis maintained by tonic contraction of adductor muscles after removal of the noxious stimuli. Hyperthermia, light anesthesia, hyperventilation, hypocapnia, and small lung volume all increase this reflex, whereas hypoxia, hypoventilation, hypercapnia, positive intrathoracic pressure, and deep anesthesia depress it. Laryngospasm can be broken with deep anesthesia or awakening. Lung inflation by application of positive airway pressure and PEEP can reduce its incidence and severity. Infants are at particular risk due to their laryngeal hyperexcitability. Premature infants with fluid in their larynx respond with apnea and breath holding (apnea at end inspiration). Fluid pooled in the hypopharynx is usually swallowed, but these small infants may develop central apnea even without fluid in the larynx. Apnea is more prevalent with plain water than with saline. Pooled secretions need to be suctioned to prevent this induced central apnea. The genioglossus displaces the hyoid and tongue anteriorly, maintaining the pharyngeal airway. This response is depressed by alcohol, sleep, and general anesthesia. Normal phasic activity of the genioglossus is easily abolished by general anesthesia, requiring high positive airway pressure or placement of an oral airway to maintain pharyngeal patency.

The diaphragm is the most important muscle for normal ventilation. Phrenic nerve injury can occur from birth injury or inadvertent surgical trauma. Unilateral phrenic nerve palsy results in paradoxical diaphragmatic movement of the paralyzed side and profound respiratory failure in young infants. Older infants have much more reserve and can maintain normal ventilation. Bilateral phrenic nerve palsy in infants or children universally requires assisted ventilation.

Ventilation may be quantified by measuring the exhaled tidal volume (V_T) over the course of a minute (V_E). Mathematically, this is

$$V_E = V_T \times f,$$

where f is the number of breaths per minute. Both f and V_T are adjusted to minimize the work of breathing. With respiratory disease that causes increased airway resistance or decreased pulmonary compliance, infants will increase their respiratory rates to more than twice normal because it requires less work to increase rate of breathing than to maintain or increase the tidal volume. The anatomical dead space may be estimated by solving the equation

$$V_D/V_T = (PaCO_2 - P_{ET}CO_2)/PaCO_2,$$

where $P_{ET}CO_2$ is the end-tidal carbon dioxide (CO_2) measured from the end-exhaled gas.

PULMONARY CIRCULATION

The pulmonary circulation and diffusion of gases across the alveolar-capillary membrane are important determinants of pulmonary gas exchange. The alveolar capillary distance is 0.3 mm. Carbon dioxide is 20 times more diffusible than oxygen. Therefore, problems with carbon dioxide diffusion do not become apparent until there is severe disease. Decreased diffusion capacity is also seen in conditions with diminished pulmonary blood flow. In utero, the normally low PaO_2 causes pulmonary vasoconstriction that results in markedly increased pulmonary vascular resistance. In the normal fetal circulation, both the right and left ventricles pump blood to the systemic circulation. Approximately 5% of the fetal cardiac output goes to the lungs, as opposed to 50% postnatally. Most of the blood pumped out of the right ventricle passes through the main pulmonary artery and through the ductus arteriosus to the descending aorta. With the newborn's first breaths of room air, the PaO_2 rises, dilating the pulmonary vasculature and increasing pulmonary blood flow. In addition, the ductus arteriosus constricts, for the first time separating the pulmonary and systemic circulations. Postnatally, $PaCO_2$ (partial pressure exerted by carbon dioxide dissolved in arterial plasma and red blood cell water) and PaO_2 still have an effect on pulmonary circulation. Elevated PaO_2 and low $PaCO_2$ dilate the pulmonary vasculature and simultaneously constrict the systemic circulation. Low PaO_2 and elevated $PaCO_2$ have the opposite effects, constricting the pulmonary vasculature, inducing pulmonary hypertension, and dilating the systemic circulation.

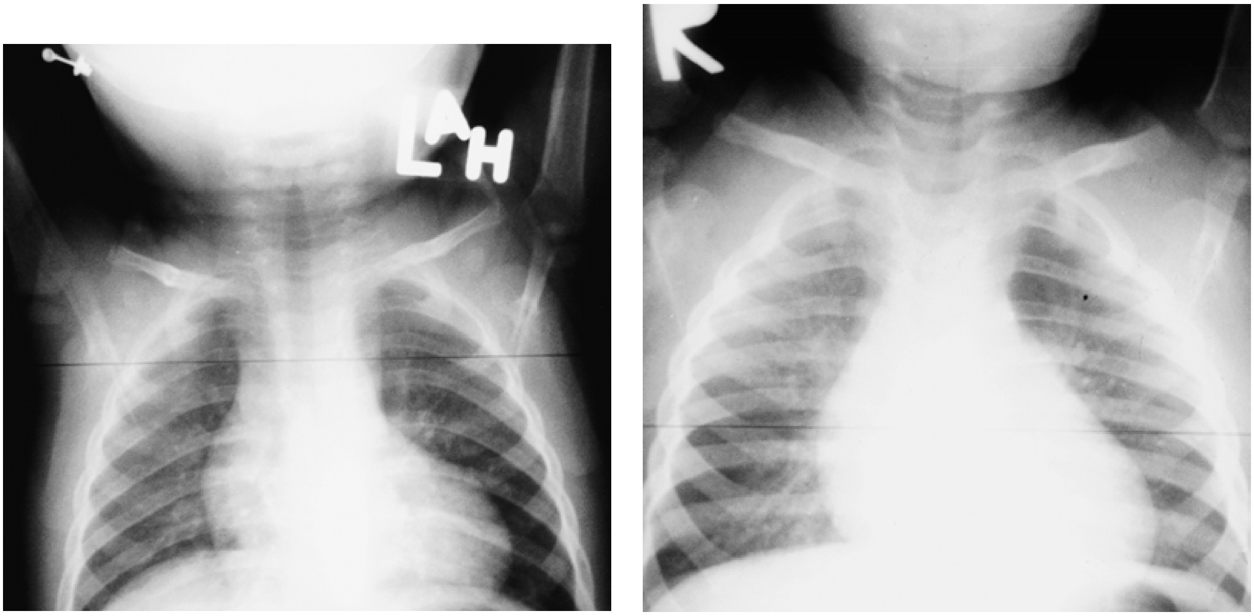
Arterial oxygen content is determined by hemoglobin concentration and arterial hemoglobin oxygen saturation. Hemoglobin oxygen saturation has many determinants such as arterial pH, body temperature, presence of 2,3-diphosphoglycerate (DPG), type of hemoglobin, presence of carbon monoxide, and oxidative state of the iron molecule in the hemoglobin. The ability of a particular hemoglobin to bind oxygen is called its p50, or the PaO_2 where hemoglobin is 50% saturated with oxygen. The p50 of adult hemoglobin is 27, for an infant 30, and a newborn 19. Newborns have high concentrations of fetal hemoglobin that bind oxygen well but decrease oxygen release at the tissue level. The normal newborn has a relatively high hemoglobin level, mean 17 g/dL. This is likely due to high fetal levels of erythropoietin stimulated by the low PaO_2 in utero and the high oxygen-binding affinity of fetal hemoglobin. By 2 to 3 months of age, infants develop a physiological anemia with a mean hemoglobin concentration of 12 g/dL. By 6 months of age, the hemoglobin concentration reaches 12.7 g/dL and remains there until the child goes through puberty, at which point adult levels are realized.

The decision to transfuse an infant prior to surgery depends on the degree and duration of anemia, intravascular volume, extent of the proposed surgery, and probability of massive blood loss. Impairment of cardiorespiratory function would also lead to transfusion at least to a normal hemoglobin concentration for age. A hematocrit greater than 25% is probably acceptable for infants older than 3 months of age without any cardiorespiratory compromise. Halothane has a minimal effect on the p50 of hemoglobin, whereas nitrous oxide causes a reversible increase in hemoglobin affinity for oxygen, decreasing oxygen unloading at the tissue level. Alkalosis has a similar effect. Because hypoxemia is quite common in infants postoperatively, pulse oximetry should be monitored and supplemental oxygen administered when oxygen saturations fall below 95%.

Ciliated cells are located in the trachea and bronchi. Serous and mucus-producing cells are located in the submucosa. The cilia beat in a serous layer covering the epithelial line. There is a discontinuous layer of mucus above the serous layer. Ciliary function is to remove mucoid secretions and foreign particles and debris. Cilia beat at an incredible 600 to 1300 beats per minute and can move mucus 1.5 to 2.0 cm per minute. Viral infections reduce ciliary motion by as much as 50%. Repeated infections can destroy cilia entirely. Breathing warm, humidified air maintains cilia. Breathing dry air or oxygen for as little as 3 hours results in complete cessation of ciliary movement. Inhaled anesthetics, 100% oxygen, and positive-pressure breaths all diminish ciliary function.

INFECTIOUS DISEASES

Infectious disease can compromise respiration at all levels of the respiratory system. Though less common in this age of antibiotics, abscess formation in the retropharyngeal



A B

Figure 16–1 ***(A)*** *This x-ray image shows the tracheal air column of a normal infant.* ***(B)*** *This x-ray image demonstrates a narrowing of an infant's tracheal air column due to edema of the trachea during an episode of croup (the "steeple" sign).*

or peritonsillar space can cause significant symptoms. Fever, dysphagia, drooling, neck rigidity, and noisy breathing are symptoms of a retropharyngeal abscess that usually occurs in children less than 5 years of age. Peritonsillar abscess occurs in older children. Symptoms include trismus, throat pain, and a muffled voice. Group A *Streptococcus* is the most common organism in both disorders. Some *Bacteroides* species also may be involved. Surgical drainage, appropriate antibiotics, and special attention to the airway are the necessary therapeutic interventions.

The next group of infectious disorders falls under the category of croup, characterized by inspiratory stridor. Supraglottitis or epiglottitis has become extremely rare due to immunization against the major causative organism; that is, *Haemophilus influenzae* type B. *Streptococcus* and *Staphylococcus* may also be causative organisms. Children with these disorders have high fever and are extremely anxious, usually sitting up with their neck extended. Diagnosis is made by visualization of a swollen and inflamed epiglottis. Antibiotics and placement of an emergent airway, preferably a nasotracheal tube, are therapeutic and life saving. The airway may be removed when there is an audible air leak around the endotracheal tube, indicating that the edema has resolved.

Laryngitis is due to viral infection and results in edema of the cords, with the resulting coarse, raspy voice. Therapy is symptomatic. However, inflammation of the subglottic space is common and potentially life threatening. The entire laryngotracheobronchial tree is usually affected. As discussed previously, small amounts of edema can cause major narrowing of the airway with a marked increase in resistance to airflow. Children present with the respiratory symptoms of a high-pitched, barky cough and inspiratory stridor. Radiographs of the neck may show a "steeple" sign, signifying edema of the trachea and subglottic space (**Fig. 16–1A, B**). Parainfluenza type 3, influenza A and B, adenovirus, respiratory syncytial virus, and Echoviruses are the usual pathogens in infants. *Mycoplasma pneumoniae* has been isolated from older children. Therapy is symptomatic; however, a short course of corticosteroids along with inhaled racemic epinephrine usually helps resolve the symptoms. Heliox helium and oxygen mixtures may be used to decrease the work of breathing, but some infants will still need endotracheal intubation if they become exhausted from the increased work of breathing through a markedly narrowed airway. Spasmatic croup occurs in children from 6 months to 1 year of age, with recurrent acute attacks of stridor in the evening or night. Symptoms resolve with patience and mist.

Bacterial tracheitis presents with symptoms similar to viral laryngotracheobronchitis, which progress to severe respiratory distress with high fever. *Staphylococcus aureus* is the most frequent organism isolated, but *H. influenzae, Moraxella catarrhalis, and Streptococcus pneumoniae* have been implicated. Routine childhood vaccinations have dramatically decreased the incidence of *H. influenzae*. Thick, purulent secretions are the cause of the respiratory distress. Bacterial superinfection of viral laryngotracheobronchitis is the likely etiology. Antibiotics are chosen to cover the usual organisms, and endotracheal intubation is often required until secretions thin and there is audible air leak with resolution of the acute infection.

Pneumonia will also present with respiratory distress due to decreased lung compliance and an increase in bronchial secretions. Children with pneumonia are frequently cyanotic as well. Most pneumonias that affect children are of viral origin. Respiratory syncytial virus (RSV), parainfluenza, adenovirus, rhinovirus, and rubeola are the common viral etiologies. Ribavirin has activity against RSV but has not improved outcome of infants requiring assisted ventilation. Acyclovir is indicated for varicella pneumonia that will likely become as rare as acute epiglottitis with the institution of universal chicken pox vaccination. The usual respiratory bacterial pathogens that cause bacterial tracheitis may also cause pneumonia, and appropriate antibiotics should be administered. *Chlamydia trachomatis* produces pneumonia in young infants born to mothers whose genitourinary tract is infected with the same organism. Half of these patients also present with significant conjunctivitis. *M. pneumoniae* causes lower respiratory infection in children older than 5 years of age. *Pneumocystis carinii* produces pneumonia characterized by severe hypoxemia in infants and children who are immunosuppressed from chemotherapy, malnutrition, or human immunodeficiency virus (HIV) disease. Appropriate antibiotic therapy has been very successful in treating this devastating disorder.

TRACHEOTOMY

A tracheotomy can be a life-saving procedure in pediatric patients with an upper airway obstruction or can be placed in patients who require prolonged assisted ventilation. With tracheotomy comes loss of olfaction and normal phonation. Patients can learn to speak by placing a one-way valve on the tracheotomy tube, allowing the exhaled breath to flow through the larynx, producing sound. With air no longer passing through the nasal airway, there is loss of air conditioning (i.e., warming, humidification, and filtering of the inspired air). There is a beneficial decrease in airway resistance unless the cannula is small, in which case there will be increased airway resistance. There is also a reduction in anatomical dead space, up to 100 mL in an adult.

Placement of the tracheostomy tube will result in immediate relief of airway obstruction; however, this may be accompanied by the development of severe respiratory failure from postobstructive pulmonary edema (POPE). Patients who develop POPE may need positive-pressure ventilation, PEEP, and diuresis until the respiratory distress resolves. The etiology of POPE is not entirely clear, but it appears to be related to marked negative pleural pressure that is generated during airway obstruction, pulling fluid into the thoracic space along with increased capillary-alveolar transmural pressure and increased capillary permeability. Also, breathing 100% oxygen during airway obstruction may lead to an increase in capillary permeability.

Air warming and humidification are functions performed poorly by the trachea and bronchi alone. Cool, dry air causes decreased mucus flow, making children with tracheotomy liable to have thickened secretions and mucosal metaplasia. Inspired air needs to be humidified to prevent atelectasis. These children are also prone to infection due to loss of filtering from the nasal airway. Humidification clearly is associated with less morbidity. After tracheotomy, the cough mechanism is less effective. Glottis closure during a cough results in compression of the tracheobronchial tree that results in peak air flows of 10 L per second upon release of glottis closure. With tracheotomy, glottis closure is not possible, making the cough ineffective.

A tracheostomy tube may be removed once the patient has an adequate upper airway, the larynx functions to protect the lower airway from aspiration, and the patient no longer requires prolonged ventilation. To ensure that the patient is ready for decannulation, the tracheobronchial tree must be free of infection and thickened secretions. There also should be no auscultatory signs of lower respiratory infection—rales, rhonchi, and wheezes. The patient must be able to breathe comfortably with the tracheotomy tube occluded, making it necessary to place a small enough tracheotomy tube that does not occlude the trachea. Pulmonary function should be evaluated for a normal tidal and respiratory rate for age, along with a negative inspiratory force of at least 20 cm H_2O to ensure adequate respiratory muscle function. The airway must be visualized endoscopically for evidence of granulation tissue at the site of the tracheostoma, other lesions in the trachea, subglottic stenosis, and vocal cord function. Suprastomal granulomas are a common sequela following tracheotomy in children. If present, the granuloma must be dissected out prior to decannulation.

DECANNULATION

Once physical dependence has resolved, the patient may be decannulated. Psychological dependence is quite common. The child and family need to be continually prepared for decannulation from the time the tracheotomy is performed. Decannulation is performed in the hospital within a few days of the endoscopy. Staff must be prepared to replace the tracheotomy tube or place an endotracheal tube if the child develops respiratory distress.

Decannulation failure is likely due to loss of laryngeal reflexes, assuming anatomical causes were ruled out by endoscopy. Airflow through the glottis is responsible for maintaining laryngeal reflexes necessary to protect the airway. Tracheotomy results in the loss of the laryngeal abductor reflex, which becomes an issue after decannulation. Without this reflex the larynx is not maintained open, and children frequently will have stridor postdecannulation until this reflex is reestablished. This problem may be ameliorated if the tracheotomy tube size is gradually made smaller over a few days, allowing for airflow through the glottis prior to decannulation, with the expectation that the laryngeal abductor reflex will reestablish prior to final decannulation. The laryngeal adductor reflex is also lost without laryngeal airflow. This results in a loss of glottis closure during swallowing and may increase the possibility of pulmonary aspiration despite the presence of the tracheotomy tube. Both of these reflexes may not return immediately after decannulation, leaving the child at risk for aspiration and stridor.

Suggested Readings

De Vries PA, De Vries CR. Embryology and development. In: Gans SL, ed. Othersen: The Pediatric Airway. Philadelphia: WB Saunders; 1991:3–16

Helfaer MA, Nichols DG, Rogers MC. Developmental physiology of the respiratory system. In: Rogers MC, ed. Textbook of Pediatric Intensive Care. Baltimore: Williams & Wilkins; 1996:97–126

Johnson JT, Reilly JS, Mallory GB Jr. Decannulation. In: Myers EM, Stool SE, Johnson JT, eds. Tracheostomy. New York: Churchill Livingstone; 1985:201–210

Motoyama EK. Respiratory physiology in infants and children. In: Motoyama EK, Davis PJ, eds. Smith's Anesthesia for Infants and Children. St. Louis: Mosby; 1996:11–67

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Normal fetal partial pressure exerted by oxygen dissolved in arterial plasma and red blood cell water (PaO_2) causes which of the following?
   A. Rhythmic fetal breathing
   B. Pulmonary vasoconstriction
   C. Constriction of the ductus arteriosus
   D. Systemic arteriolar constriction
2. In premature infants, hypoxia has which of the following effects?
   A. Tachypnea
   B. Apnea
   C. Hyperventilation
   D. Decreased frequency of periodic breathing
3. Helium and oxygen mixtures (heliox) will decrease the work of breath because of which of the following reasons?
   A. Heliox decreases airflow resistance in areas of laminar flow.
   B. Helium has higher density than nitrogen.
   C. Helium increases hemoglobin oxygen saturation.
   D. Heliox decreases airflow resistance in areas of turbulent airflow.
4. Which of the following is an infectious etiology of laryngotracheobronchitis, or "croup"?
   A. *Haemophilus influenzae*
   B. *Staphylococcus aureus*
   C. *Branhamella catarrhalis*
   D. Parainfluenza type 3
   E. Varicella
5. The narrowest portion of an infant's upper airway is
   A. The carina
   B. The vocal cords
   C. The subglottic space
   D. The nasal airway
   E. The pharynx

Chapter 17
BRANCHIAL CLEFT ANATOMY AND CONGENITAL NECK MASSES

Gerald B. Healy

Congenital defects may arise as a result of disturbances of the orderly embryological development of structures originating from the primitive branchial apparatus. The branchial apparatus develops during the third and fourth fetal weeks and persists until the end of the sixth week of fetal development. The branchial arches, which form in the cephalic region, consist of five parallel bars of mesoderm from which cartilage, bone, and muscle form. Each of these bars has its own nerve and blood supply. The branchial arches are separated externally by branchial clefts or grooves consisting of ectoderm. They are separated internally by branchial or pharyngeal pouches lined by entoderm. Each of the clefts and pouches differentiates into various anatomical structures.

Branchial cleft anomalies represent disturbances in the mechanics of precise embryological development in the head and neck region. A clear understanding of the developmental anatomy is necessary to understand the clinical implications of disorders of development in the neck.

DEVELOPMENTAL ANATOMY

The branchial apparatus appears to undergo its major development and differentiation between the third and seventh embryological weeks. This apparatus is composed of a group of five mesodermic arches that are paired and separated by four pairs of invaginations of ectoderm and entoderm referred to respectively as branchial cleft pouches. Each of these arches is supplied by its own artery and nerve and develops into well-defined structures of skeletal muscle and connective tissue in the fully formed fetus. Each of the clefts and pouches differentiates into various anatomical structures, as noted in **Table 17–1.** Over 90% of the branchial cleft defects in the human arise from the second pouch and groove and consequently are said to pass between the internal and external carotid arteries. Several theories exist as to how branchial cysts and fistulas are formed. The most widely accepted theory is that these fistulas are derived from remnants of branchial grooves and pouches that have a failure of either fusion or burying of cell rests of the branchial grooves.

Knowledge of the vascular supply of each arch is critical in understanding the implications of distortions of vascular development in the newborn. Each arch is supplied by a central artery, which then connects the paired primitive dorsal and ventral aortas. The first arch artery probably results in the development of the facial branch of the external carotid artery. The artery of the second

TABLE 17–1 STRUCTURES THAT ARISE FROM DIFFERENTIATION OF BRACHIAL ARCHES

Arch	Nerve	Structure	Muscle	Artery
First	Trigeminal	Mandible, body of incus, head and neck of malleus, major salivary glands, tympanic membrane, eustachian tube	Masticator muscles, tensor tympani, anterior belly of digastric	Facial
Second	Facial (VII), cochleovestibular (VIII)	Two thirds of long process of incus, manubrium of malleus, crura and head of stapes, styloid, lesser cornu of hyoid, upper body of hyoid, tonsil	Platysma, muscles of facial expression, stapedius, posterior belly of digastric	Stapedial
Third	Glossopharyngeal (IX)	Greater cornu of hyoid, thymus, inferior parathyroid, body of hyoid	Superior constrictors	Internal carotid
Fourth	Vagus (X)	Thyroid cartilage, epiglottis, superior parathyroid	Inferior constrictor, laryngeal musculature	Aortic arch, right subclavian
Fifth	Spinal accessory (XI)	Arytenoid cartilage, cricoid cartilage, lungs	Portion of laryngeal musculature	Pulmonary ductus

arch becomes vestigial but may persist as the stapedial artery on rare occasions. The arteries of the third arch fuse at the primitive cephalad extension of the dorsal aortas to form the internal carotid arteries, and the caudal ends of the dorsal aortas fuse to become the descending aorta. The ventral aortas give rise to the common and external carotid arteries. The right fourth artery becomes the subclavian artery and the left fourth becomes the aortic arch. The left fifth becomes the pulmonary artery. All other branches most likely disappear.

The nerves supplied to each of the five arches are as follows: first arch, V; second arch, VII and VIII; third arch, IX; fourth arch, X; and fifth arch, XI.

CLASSIFICATION

Numerous terms are used to describe the clinical findings in disorders of branchial (cervical) development. A branchial cyst refers to a mucosal or epithelial lined structure with no external openings. These cysts may contain respiratory epithelium as well. There are considerable amounts of subepithelial lymphoid tissue, and sebaceous glands and salivary tissue may also be present. These cysts are usually filled with a viscid fluid.

A branchial sinus refers to a tract, with or without a cyst, that communicates with the skin. These tracts may be lined by pseudostratified columnar epithelium, and lymphoid tissue may be present.

A branchial fistula refers to a tract that communicates with the pharyngeal or hypopharyngeal structures. The opening in the pharynx may extend to either a cystic structure in the neck or directly to an external opening in the skin.

CLINICAL PRESENTATION

Branchial cleft cysts represent the most common noninflammatory lateral neck masses in children. These anomalies most commonly originate from the first, second, and third branchial apparatus. Fourth branchial anomalies are extremely rare.

Anomalies of the first branchial apparatus are uncommon. They make up less than 1% of all branchiogenic abnormalities. These anomalies are usually related to the auricle and surrounding face and should be considered to be distinct from preauricular cysts and sinuses that result from failure of fusion of the aural folds of the first branchial arch. They are subdivided into two types. Type I defects, which contain epidermoid elements only, are lesions that usually present as duplications of the external auditory canal and thus are usually found close to that structure. There may be an opening in the skin above the hyoid bone, which signifies a sinus tract that runs in close association with the parotid gland as well as the branches of the facial nerve. The sinus tract frequently terminates deep in the external auditory canal or even in the middle ear space.

Type II anomalies are seen more frequently. These contain both ectodermal and mesodermal elements. An external opening may be present leading to a cyst or sinus tract that courses through the parotid gland, usually passing closer to the facial nerve than type I anomalies. The tract frequently terminates at the junction of the bony and cartilaginous portions of the external auditory canal. Patients frequently present with aural discharge that is unresponsive to traditional treatment.

Lesions of the second branchial apparatus are the most common. These can present as a cyst, sinus tract, or fistula. Second branchial cysts usually present anterior to the sternocleidomastoid muscle and have an associated tract that passes deep to the external carotid artery and lateral to the glossopharyngeal and hypoglossal nerves (**Fig. 17–1**). The tract from this lesion frequently ends in the tonsillar fossa. These lesions present as painless masses below the angle of the mandible and anterior to the anterior border of the sternocleidomastoid muscle (**Fig. 17–2**).

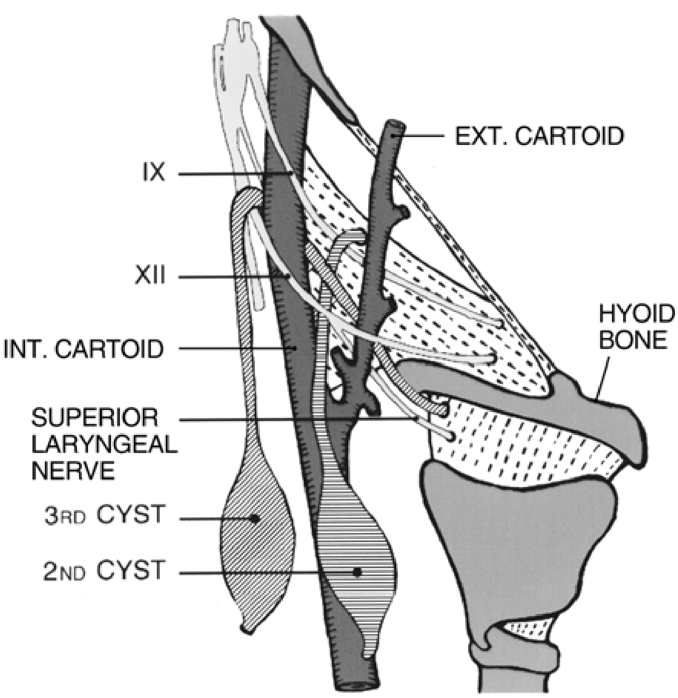


Figure 17–1 *Relationship of second and third branchial cleft anomalies to the carotid artery and cranial nerves.*

Figure 17–2 *Classic second branchial cleft fistula.*

Abnormalities of the third branchial apparatus are uncommon (**Fig. 17–3**). Cysts associated with this anomaly are usually low in the neck and anterior to the sternocleidomastoid muscle. There is frequently an external opening. The fistula associated with this defect runs deep to the carotid artery, lateral to the hypoglossal

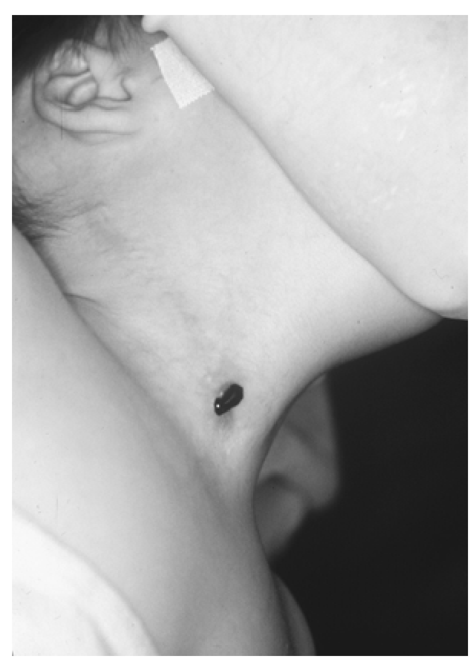


Figure 17–3 *External opening of third branchial cleft sinus.*

nerve, but superficial to the vagus nerve, which is a fourth arch derivative. The fistula usually enters the pharynx at the level of the pyriform sinus (**Fig. 17–1**).

Defects of the fourth branchial apparatus are extremely rare, if they exist at all. Theoretically, these anomalies should present with a cystic structure actually located in the chest. There may be an external opening along the lower, anterior border of the sternocleidomastoid muscle. The fistula associated with this lesion should pass under the arch of the aorta on the left side or the subclavian artery on the right. These structures are fourth branchial derivatives. The fistula would then theoretically ascend in the neck and have an internal opening in the esophagus, a fourth branchial derivative.

DEFINITIVE TREATMENT

The definitive treatment for all of these lesions is surgical excision. These lesions may be quite troublesome for patients, especially if there is an external opening to the skin. There are usually persistent secretions of mucus, and once infection has occurred, there is a high tendency for recurrence. Carcinomatous degeneration of branchial cysts has been reported to occur on extremely rare occasions in the adult population.

OTHER CONGENITAL MASSES

Other congenital masses may occur in the head and neck region of children. Perhaps the most common, after branchial cleft anomalies, is the thyroglossal duct cyst. Embryologically, the primitive thyroid begins its development in the pharynx in the area of the foramen caecum of the tongue base. The duct descends through the anterior neck, reaching the level of the larynx. During this descent, the duct passes close to or in fact through the central portion of the hyoid. This duct usually disappears before birth, but in a small number of individuals, the duct persists potentially anywhere along its path (**Fig. 17–4**).

These lesions may become infected on a repeated basis, thus prompting surgical excision, which is the definitive treatment. Although rare, carcinomatous degeneration of these lesions occurs, which is also a valid reason for removal.

Before removal, the presence of normal thyroid tissue should be documented by the use of ultrasound or thyroid scan. In extremely rare situations, the patient's only functioning thyroid may be located within the cyst itself. These diagnostic studies will make this determination prior to surgery.

Dermoids usually occur in the midline of the neck in a similar location to the thyroglossal duct cyst. These lesions are frequently located in the submental region and are well encapsulated, with no accompanying sinus tract.

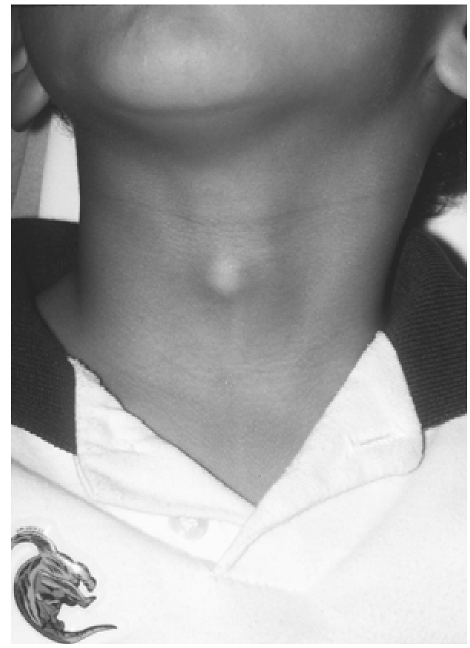


Figure 17–4 *Classic anterior thyroglossal duct cyst.*

Cervical thymic cysts may present in the midline or lateral cervical region. These lesions arise in the third branchial pouch as a result of the migration of the thymus from the upper neck to its final location below the clavicles. Thymic remnants may persist as cords or cysts along this pathway.

VASCULAR MALFORMATIONS

The most common lesions in this group are lymphatic malformations and hemangiomas. Lymphatic

Figure 17–5 *Large hemangioma of the facial region in a young infant.*

malformations are multiloculated, painless masses that usually present shortly after birth and are far less distinct than congenital cysts of the neck. These lesions may enlarge to become cosmetically disfiguring. Treatment is by surgical excision; however, these lesions do not grow along defined tissue planes and may envelop vital structures. This must be strongly considered when surgical planning is undertaken.

Hemangiomas are said to represent the most common neoplasms in children. They usually present in the facial area during the first few months of life and have their greatest growth rate during the first 12 months (**Fig. 17–5**). In most cases, spontaneous involution occurs. In rare cases where exceedingly large lesions are present, the cardiovascular system can be adversely stressed. In these cases, medical therapy in the form of steroids or interferon may be indicated. Surgical resection is rarely necessary.

SUMMARY

These disorders of embryological development represent important clinical entities. A clear understanding of these lesions is necessary for the clinician treating medical and surgical diseases of the head and neck in children and adults.

Suggested Readings

Albers GD. Branchial anomalies. JAMA 1963;183:399–409.

Chandler JR, Mitchell B. Branchial cleft cysts, sinuses and fistulas. Otolaryngolog Clin North Am 1981;14(1):175–186.

Gage JF, Lipman SP, Myers EN. Diagnosis and Management of Congenital Head and Neck Masses [self-instructional package]. American Academy of Ophthalmology and Otolaryngology; 1976.

Proctor B, Proctor C. Congenital lesions of the head and neck. Otolaryngolog Clin North Am 1970;3(2):221–248.

Work WP. Newer concepts of first branchial cleft defects. Laryngoscope 1972;82:1581–1593.

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. The stapedius muscle is innervated by cranial nerve
   - A. V
   - B. VI
   - C. VII
   - D. IX
   - E. X
2. A patient is noted to have a draining sinus tract deep in the external canal. Surgical resection will include a dissection of
   - A. The jugular vein
   - B. The superficial temporal artery
   - C. The hyoid bone
   - D. The stapes
   - E. The facial nerve
3. The central portion of the hyoid bone should be removed when removing a
   - A. Dermoid
   - B. Thyroglossal duct cyst
   - C. Thyroid nodule
   - D. Branchial cleft cyst
   - E. Thymic cyst
4. The removal of a second branchial cleft sinus tract may require removal of
   - A. The thyroid gland
   - B. The tonsil
   - C. The parotid gland
   - D. The hypoglossal nerve
   - E. The hyoid bone

Chapter 18

PATHOPHYSIOLOGY OF STRIDOR AND AIRWAY DISEASE

John H. Greinwald and Robin T. Cotton

The management of airway obstruction in children is considered a vitally important aspect of residency training and our practice of otolaryngology. Because of the often immediate consequences in pediatric airway disease, a clear understanding of the pathophysiology, diagnostic evaluation, and treatment options should be understood and practiced. The development of fiberoptic technology has greatly improved our diagnostic abilities, while maintaining a margin of safety in difficult airway conditions. Therefore, with a broad knowledge of the etiologies of pediatric airway disease and the application of fiberoptic evaluation, the diagnosis is often readily apparent prior to any operative intervention. This chapter will review the physiology of stridor and airway noise, the diagnostic evaluation, and the management options in patients with pediatric airway obstruction.

DEFINITION OF STRIDOR AND AIRWAY NOISE

Stridor is a term that is often used to describe any abnormal airway noise. This broad definition, unfortunately, is supported by many textbooks, which typically describe stridor as a high-pitched musical noise, often inspiratory in nature, which denotes laryngeal obstruction. Stridor may be high or low pitched, inspiratory, expiratory, or both, and may represent obstruction at nearly any part of the respiratory tract. The Latin word *stridere* means "to make a harsh noise," although in practice the noise of airway obstruction may be soft or harsh in quality. The acoustic quality of stridulous noises is typically described as harmonious in frequency and pitch, greater than 200 msec in duration and best heard over the cervical region.

PHYSIOLOGY

The mechanism of stridor is that of a narrowing of the airway at some point between the oral or nasal cavity and the distal bronchi. The actual sound of stridor is the result of abnormal airflow patterns that develop from either anatomical or functional narrowing of the airway. To better understand this phenomenon, one must understand that gas moving through a partially obstructed tube is subject to certain principles of physics. As the lumen begins to narrow, the net velocity of the air molecules must increase to allow the same volume of air to pass. According to the principles of Venturi, the force vectors will shift 90 degrees forward, causing a decrease in the lateral wall pressure, thereby inducing airway collapse of the generally compliant airway in children. As airway obstruction proceeds to near-complete narrowing, turbulence of the typically laminar airflow commences. The narrowed airway begins to have rhythmic vibrations and produces an often musical sound, not dissimilar to the production of sound from the reed of a woodwind instrument. Because of its vibratory nature, stridor can often be felt as well as heard on examination.

The small size of the pediatric airway is an important factor in the development of airway obstruction, particularly in naturally narrow areas such as the glottis and subglottis. Because of its small size and complete encirclement by the cricoid cartilage, the subglottis is particularly susceptible to narrowing. Five millimeters is considered the normal-size subglottic airway in a newborn, with progressive enlargement with age (**Table 18–1**). Cross-sectional area of the airway can be calculated by the equation area $= \pi r^2$, where $r =$ radius. Because of the exponential relationship between the area of a circle and the radius, small changes in the radius can account for large changes in airway cross-sectional area. A 20% reduction in the radius of the airway; that is, 1 mm of narrowing in a normal neonate, would account for a 36% reduction in cross-sectional area. Two millimeters of narrowing (40% of the radius) would contribute to a 64% reduction in the airway area. Children, particularly neonates and infants, deserve vigilant attention to any airway complaints due to the narrow margin of safety their small airways afford.

Stridor is often incorrectly used to describe other pathological airway noises such as stertor, wheezing, rhonchi, and rales. Location and the musical versus nonmusical quality of the sound usually delineate the proper diagnosis. Another musical noise that is similar to stridor is a wheeze. Although acoustically and pathophysiologically similar to stridor, wheezing is best heard over the chest, is often bilateral, and responds to bronchodilator medications. Stertor is a nonmusical noise described as an acoustically heterogeneous, low-pitched reverberation of the oropharyngeal tissues and better known as snoring. Rhonchi are similarly low-pitched noises of the large airways of the chest typically caused by secretions. Finally, rales (or "crackles") are a fine, nonmusical, popping-like noise heard in the peripheral lung fields lasting less than 200 msec. The sound of rales represents the sudden opening of collapsed alveoli, usually indicative of the presence of fluid.

TABLE 18–1 NORMAL AIRWAY SIZE BY AGE

Age	Normal Subglottic Airway Diameter (mm)	Expected Endotracheal Tube Size	Expected Bronchoscope Size
Premature	3.5–4.5	2.5–3.0	2.5
0–3 months	5.0	3.5	3.0
3–9 months	5.5	4.0	3.5
9–24 months	6.0	4.5	4.0
2–4 years	6.5–7.0	5.0	4.0
4–6 years	7.5	5.5	5.0
6–8 years	8.0	6.0	6.0

LOCALIZATION

Any child with airway obstruction should be carefully assessed from nasal cavity to distal lung to determine the cause of the airway noise.

During the examination, careful observation and auscultation of the patient can properly localize the point of airway obstruction prior to any endoscopic evaluation. As previously mentioned, nonstridulous noises can often be identified and properly diagnosed. The otolaryngologist is assisted by dividing the airway into acoustic zones: (1) naso- and oropharynx; (2) supraglottis; (3) glottis, subglottis, and extrathoracic trachea; (4) intrathoracic trachea; and (5) distal pulmonary airways. Naso- and oropharyngeal obstruction typically produces stertorous noises and is generally not life-threatening. One exception is that of bilateral choanal stenosis in neonates due to their obligate nasal breathing requirement. Likewise, distal pulmonary airway noises such as wheezing and rales are easily determined on auscultation of the lung fields.

Significant and life-threatening airway obstruction can occur in the laryngeal and tracheal regions; therefore, a rapid and accurate assessment of the location of the airway lesion is imperative. Although factors such as the time course, severity, and presence of cyanosis are helpful to formulate the diagnostic and treatment plans, this information adds little to the localization of the lesion. Localization of the airway lesion can be accomplished by evaluating four factors: retractions, stridor, voice, and feeding [use the mnemonic "Real stridor is very frightening" (RSVF)] (**Table 18–2**).

The supraglottic airway zone is characterized by relatively loosely supporting structures, which, during inspiration, tend to collapse as a result of the Venturi

TABLE 18–2 STRIDOR LOCALIZATION BY ANATOMIC SITE

	Retractions	Stridor[1]	Voice	Feeding
Naso/oropharynx	Minimal[2]	Stertor[2]	Normal	Normal[2]
Supraglottis	Marked and severe	Inspiratory and high pitched	Muffled	Abnormal
Glottis/subglottis	Mild to severe	Biphasic and intermediate pitched	Normal to very abnormal (barking cough)	Normal
Intrathoracic trachea	Mild to severe	Expiratory and low pitched	Normal (seal-like cough)	Normal

[1]The quality of the airway noise.
[2]Unless associated with complete nasal obstruction in a neonate.

effect. This collapse can be made particularly worse during periods of air hunger. The inspiratory stridor of supraglottic obstruction is high pitched due to the vibratory ability of the loose supraglottic tissues. During expiration, the tissues open, and airflow is unimpeded. The loose tissues surrounding the supraglottic region are prone to retract during periods of airway obstruction because of its cervical (extrathoracic) location. Supraclavicular and thoracic retractions are common even with minimal stridor present. Due to the high compliance of the infant rib cage, retractions can be severe. The supraglottic region primarily contributes to the resonance of the voice; therefore, obstruction typically produces a muffled vocal quality. Because of its collocation with the pharynx, supraglottic obstruction is often associated with feeding problems. This may range from aspiration to the inability to swallow, with significant drooling.

The third zone consists of the glottis, subglottis, and extrathoracic trachea. These areas share a similar characteristic in that each is a relatively rigid and noncollapsible tube. The mucosa is tightly bound to the rigid supporting structures of the vocal ligament, cricoid cartilage, and upper tracheal rings. The airflow is less regulated by fluid dynamics, because there is little compliance of the airway tissues in this region, than by the absolute cross-sectional area. Critical airway narrowing is heralded by the onset of biphasic stridor, occurring in both inspiration and expiration. The stridor is often severe and associated with tremendous respiratory effort, with both supraclavicular and intercostal retractions. The less compliant vibratory tissue in this region also contributes to the intermediate pitch of the stridor, as compared with the high-pitched stridor of supraglottic obstruction. With glottic obstruction, the voice is obviously affected, and the patient may be hoarse or completely aphonic. Subglottic obstruction often displays a normal voice because the free edge of the vocal cords is not involved, although a characteristic "doglike" barking cough is appreciated in patients with subglottic narrowing. In general, there are no direct feeding or swallowing problems, except those derived from being dyspneic. One exception to this rule is the possibility of aspiration with a vocal fold paralysis.

The intrathoracic trachea is dynamically quite unique. During inspiration, the negative intrathoracic pressure tends to maintain the integrity of the tracheobronchial lumen. It is during expiration that the relative positive intrathoracic pressure may contribute to dynamic airway collapse. This may be seen routinely during bronchoscopy because the posterior membranous wall tends to collapse during expiration. Of course, the lumen is maintained by the rigidity of the tracheal rings. Endobronchial lesions or foreign bodies only enhance the airway collapse during expiration. The expiratory stridor usually maintains its musical quality but is harsh, similar to a distal airway wheeze. Retractions are typically not present until airway collapse is imminent. Voice and feeding are typically normal, although, again, a characteristic "seal-like" barking cough may be present with tracheal obstruction.

DIAGNOSTIC APPROACH TO STRIDOR

The clinician should inquire about the age of onset, initiating or relieving factors, and overall health of the child. Stridor at birth typically represents a fixed anatomical lesion, whereas dynamic conditions such as laryngomalacia usually present after several weeks of life. Progressive stridor often denotes either a growing tumor (subglottic hemangioma) or scar deposition (subglottic stenosis).

Positional changes to the stridor are characteristic of a functional supraglottic disorder, such as laryngomalacia (worsening stridor in the supine position and improvement in the prone position). Disorders of the cardiovascular, gastrointestinal, and neurological system may either mimic primary airway disorders or contribute to their severity. For instance, the diagnoses of cyanotic heart disease, vascular anomalies, gastroesophageal reflux, and central apnea should be carefully evaluated. For neonatal patients, perinatal historical facts including prematurity, difficult labor and delivery, and postnatal intubations should be carefully reviewed. A thorough assessment, using the previously mentioned RSVF mnemonic, will greatly assist in localizing the airway lesion. Pulse oximetry can be useful in the acute setting to document oxygen desaturations associated with airway compromise. A complete history and physical examination will often provide the clinician with enough information to make a presumptive diagnosis, but nothing should supplant the otolaryngologist from performing an endoscopic evaluation.

Endoscopic evaluation is the definitive manner to diagnose all airway disorders. The advances in fiberoptic technology have allowed the clinician easier access to and a clearer picture of the airway. Flexible fiberoptic laryngoscopy (FFL) is a mandatory part of the assessment in children with stridor. A kind bedside manner, gentle technique, and application of a topical anesthetic/vasoconstrictor will usually allow visualization in most children over the age of 4 years. Gentle restraint will be required for younger children. A more detailed assessment of the cause of the stridor can be ascertained because this evaluation is dynamic and is not performed with general anesthesia. The two most common causes of stridor (laryngomalacia and vocal fold dysmotility disorders) can be easily diagnosed. Further benefit is gained by the use of a camera and monitor, together with video-recording and picture-taking ability to document the findings. If no abnormalities are found on flexible examination, the clinician should be concerned about a lower airway obstruction.

Much controversy currently surrounds the necessity for formal direct laryngoscopy and bronchoscopy (DLB) conducted in the operating room for all patients with stridor. DLB allows the examination of the entire airway and both correctly identifies the diagnosis and rules out other airway lesions. In general, patients with stridor should have a DLB performed. This is particularly important for those patients who have cyanotic or "dying" spells [apparent life-threatening events (ALTEs)], require a surgical intervention to correct the stridor (no matter what the cause), have multiple congenital anomalies, or have an uncertain diagnosis. One area that remains controversial is the necessity of DLB in patients with mild laryngomalacia (diagnosed by FFL) and who have none of the aforementioned criteria necessitating a DLB. As many as 10% of patients with laryngomalacia, in general, will demonstrate a second airway lesion. Many otolaryngologists feel that second lesions are uncommon in the subset of patients with mild laryngomalacia and that if a second lesion is present, it usually does not require therapy. One example of this would be mild tracheomalacia from innominate artery compression associated with laryngomalacia. Others believe that these possible second lesions located below the vocal cords (subglottic stenosis and tracheomalacia) are important to diagnose and recommend routine DLB for all patients with laryngomalacia. No matter which philosophy you subscribe to, careful monitoring and a thorough evaluation in all children with stridor is warranted.

The radiological evaluation for children with airway noise should include an anteroposterior (AP) and lateral radiograph of the chest and neck. These views are obtained during inspiration and expiration and should be of good quality. Expiratory views are often helpful to identify air trapping if a bronchial foreign body is suspected. Due to the redundancy of the soft tissue and the compliance of the cartilaginous structures in the upper airway of children, masses seen on expiratory views should be viewed with caution. Lateral views of the upper airway obtained during deep inspiration, with careful attention to the pharyngeal and tracheal air column, retropharyngeal soft tissue width, and epiglottis can provide valuable clues to the diagnosis of airway obstruction. High-voltage AP views best show the tracheal air column; particular attention should be paid to the region of the subglottis for evidence of narrowing of the air column.

Other types of radiographic evaluations that may be used on a case-specific basis include videofluoroscopy, bronchography, barium swallow study, computed tomography (CT), magnetic resonance imaging (MRI), and ultrasonography (US). Videofluoroscopy alone has been used to detect lower airway foreign bodies, diaphragmatic immobility, and tracheomalacia. Videofluoroscopic exam during barium swallow (BaS) is very useful in detecting aspiration with or without pharyngeal dyscoordination and encroachment of the esophageal lumen by vascular anomalies. CT or MRI can confirm aberrant mediastinal vessels and tissue masses (i.e., lymphangiomas). Bronchography has limited usefulness in light of the superior evaluation now obtained by rigid or flexible fiberoptic evaluation. US has been used to complement the endoscopic evaluation for vocal cord mobility disorders.

Gastroesophageal reflux disease (GERD), better termed pharyngoesophageal reflux, has recently been appreciated as an important underlying factor in some patients with stridor. Reflex laryngospasm, vocal fold dyskinesia, reactive airways disease, bronchitis, laryngomalacia, and subglottic stenosis have all been implicated in patients with GERD. Symptoms of overt regurgitation may be absent, so diagnostic studies such as BaS, milk scans, and pH probes (preferably a double-probe study) may be necessary. It should be noted, though, that normative data for pH probe results in young children are currently lacking, making interpretation difficult. More direct techniques of diagnosing the pathological changes of GERD may prove more useful. Bronchoalveolar lavage for analysis of lipid-laden macrophages is a very sensitive test of aspiration, and distal esophageal biopsies can easily show the inflammatory changes and basal cell hyperplasia associated with GERD. These latter two studies can be done in conjunction with the routine DLB, thereby not subjecting the patient to multiple diagnostic evaluations.

Other special diagnostic tests that may be used include polysomnography, pulmonary function tests, and acoustic analysis. The first can confirm the presence of an obstructive process or diagnose central apnea; the latter two studies have limited clinical usefulness.

AIRWAY ANOMALIES

The otolaryngologist must consider a broad differential diagnosis when presented with a child with stridor and airway noise. Although laryngomalacia accounts for 60% of all cases of stridor in children, vocal fold dysmotility disorders, subglottic stenosis, tracheomalacia, and foreign bodies contribute another 30% of cases. The diagnosis in the remaining 10% of cases is from a large list of airway disorders (**Table 18–3**). As with identification of the source of the stridor by using the RSVF mnemonic, it is best to narrow the possible diagnoses of a child with an airway disorder by considering the differential diagnosis by anatomical site. Classifying the diagnosis in each anatomical site by each etiological category—congenital, inflammatory, neoplastic, iatrogenic, traumatic, and neurogenic—ensures a complete list of possible causes. The following section includes common and/or interesting etiologies of stridor and airway noise with which all otolaryngologists should be familiar.

Nasal/Pharyngeal

Choanal Stenosis/Atresia

Obliteration of the posterior choana is the result of a failure of the nasobuccal membrane to rupture during the fourth week of embryonic life. The resulting atretic plate may be only membranous, but in 90% of cases it has a bony component. Choanal atresia occurs in 1 in 5000 live births, and there is a female to male predominance of 2:1, with most lesions being unilateral and occurring on the right side. Up to 50% of cases are associated with other anomalies, predominantly with the CHARGE association. This includes colobomas of the eyes, a variety of cardiac (heart) defects, choanal atresia, retarded growth and development, genital anomalies in males, and external ear malformations with or without a mixed hearing loss. Bilateral choanal atresia occurs in ~60% of cases associated with CHARGE.

Infants with any form of bilateral nasal obstruction will universally develop cyclical airway distress because neonates are obligate nasal breathers for the first 6 to 8 weeks of life. During crying spells, infants reflexively mouth breathe, and their airway obstruction improves. Once quiet, the respiratory distress and cyanosis will recur. Unilateral nasal obstruction often presents later in infancy or childhood as rhinorrhea and congestion that is unresponsive to medical therapy. Diagnosis is made by the inability to pass catheters through the nose or by nasal endoscopy (flexible or rigid), but a CT scan is always obtained to confirm the physical findings.

Immediate treatment in the neonatal period includes the placement of an oral airway, McGovern nipple, and possibly endotracheal intubation. If possible, surgical repair is delayed until the child is older than 2 years to allow better visualization during surgery. The most popular form of therapy is removal of the atretic plate under endoscopic or microscopic visualization. Powered instrumentation (microdebriders), lasers, otologic drills, and sinus instruments all assist in the resection. Soft silicone stents are typically placed for 2 weeks to 4 months, with most patients being stented for 4 to 6 weeks.

Oral Synechiae/Persistent Buccopharyngeal Membrane

Similar to choanal atresia, the buccopharyngeal membrane that divides the oral cavity from the oropharynx can fail to divide in the fourth week of gestation, and abnormal attachments can remain. This condition is commonly associated with cleft palate and lip, microglossia, micrognathia, microstomia, synechial bands, temporomandibular joint ankylosis, and polyhydramnios.

Table 18–3 Causes of stridor and airway noise in children

Nasal/Pharyngeal

Congenital
- Nasal aperture stenosis
- Choanal atresia/stenosis
- Nasal masses (dermoid, teratoma, encephalocele)
- Oral synechiae/persistent buccopharyngeal membrane
- Oral masses (ranula, dermoid, thyroglossal duct cyst)
- Craniofacial anomalies (Pierre Robin syndrome, micrognathia, Treacher Collins syndrome)
- Lymphangioma

Inflammatory
- Adenotonsillar hypertrophy
- Deep neck abscess (retropharyngeal, parapharyngeal)
- Ludwig's angina
- Nasal polyps
- Mononucleosis

Neoplastic
- Rhabdomyosarcoma
- Teratoma
- Juvenile nasopharyngeal angiofibroma

Trauma
- Postoperative edema
- Penetrating objects

Supraglottic

Congenital
- Laryngomalacia
- Atresia
- Web
- Epiglottic cysts
- Saccular cysts
- Lymphangioma

Inflammatory
- Epiglottitis
- Abscess
- Allergy
- Neoplastic
- Papillomatosis (RRP)
- Chondroma
- Neurofibroma

Trauma
- Foreign bodies
- Laryngeal fracture
- Inhalation/caustic burns
- Postoperative edema

Glottic/Subglottic

Congenital
- Stenosis (malformations of the cricoid, i.e., elliptical shape or trapped first tracheal ring)
- Vocal cord dyskinesia (paralysis, paradoxical motion)
- Atresia
- Web
- Lymphangioma

Inflammatory
- Viral laryngotracheitis (croup)
- Bacterial laryngitis (diphtheria)
- Allergy
- Sarcoid
- Fungal (coccidiomycosis)
- Tuberculosis
- Wegener's granulomatosis

Neoplastic
- Papillomatosis (RRP)
- Hemangioma
- Granular cell myoblastoma
- Neurofibroma
- Sarcoma (rhabdomyosarcoma, fibrosarcoma, chondrosarcoma)

Traumatic
- Stenosis (fibrosis, i.e., post–prolonged intubation)
- Vocal fold paralysis (postintubation, postductus ligation)
- Laryngeal fracture

Neurogenic
- Gastroesophageal induced laryngospasm
- Vocal fold paralysis (Arnold-Chiari malformation, familial abductor paralysis)
- Tetanus
- Tetany secondary to hypocalcemia

Tracheal

Congenital
- Stenosis
- Cartilaginous
- Tracheomalacia
- Primary
- Secondary (vascular or cystic compression)
- Complete tracheal rings (segmental, funnel, complete)
- Fibrous
- Web
- Associated with tracheoesophageal fistula
- Atresia

Inflammatory
- Viral/bacterial tracheitis
- Tuberculosis
- Fungal (histoplasmosis)

Neoplastic
- Papillomatosis (RRP)
- Mucoepidermoid carcinoma
- Fibrous histiocytoma
- Leiomyoma

Traumatic
- Foreign bodies
- Posttracheotomy

Underlined diagnoses are considered common causes of stridor and airway noise.
(Modified with permission from Cotton RT, Reilly JS. Stridor and airway obstruction. In: Bluestone CD, Stool SE, Kenna MA, eds. Pediatric Otolaryngology. 3rd ed. Philadelphia: WB Saunders; 1996.)
RRP, recurrent respiratory papillamotosis.

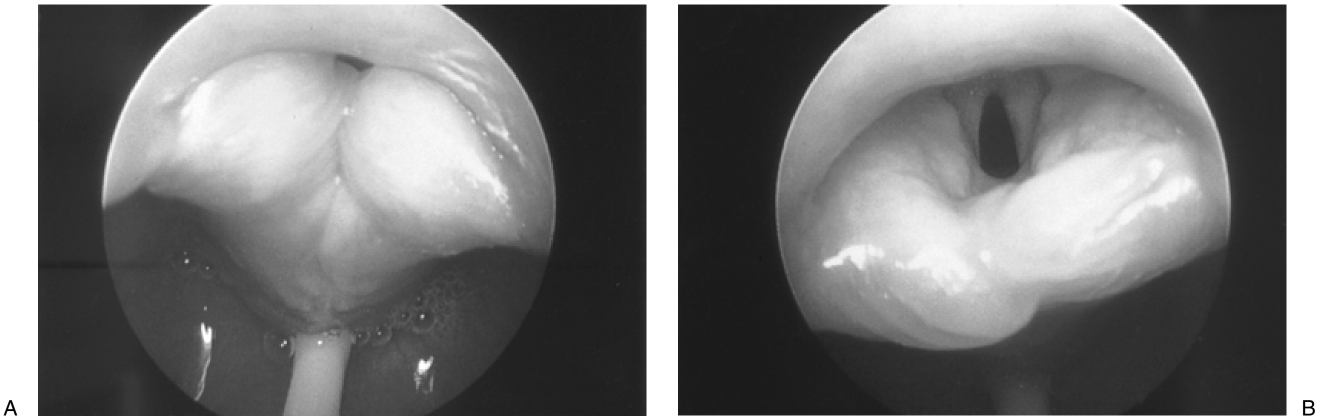


Figure 18–1 *(**A**) Inspiratory and (**B**) expiratory views of the larynx in a child with laryngomalacia. Note the prominent inspiratory collapse of the supra-arytenoid tissue, epiglottis, and arytenoepiglottic folds.*

These infants may or may not present with airway distress at birth due to adequate nasal patency. Immediate otolaryngology consultation should be obtained because, in case of the potential airway distress, oral endotracheal intubation is typically impossible. Initial treatment for children with this condition includes airway stabilization with a tracheotomy, with anesthesia delivered through mask ventilation or fiberoptic endotracheal intubation. A gastrostomy tube is often required for long-term nutritional support.

Simple division via an intraoral approach can accomplish definitive correction of isolated synechial bands. A complete buccopharyngeal membrane will require a more complex intraoral resection with mucosal flaps and long-term stenting. Revision surgery is often required in patients with complete buccopharyngeal membranes. CT evaluation, particularly with three-dimensional reconstruction, is helpful to evaluate the craniofacial skeletal deformity. Mandibular advancement or distraction osteogenesis should be considered when treating mandibular hypoplasia. The timing of mandibular surgery is controversial, with most surgeons delaying definitive repair until at least 5 years of age.

Supraglottic

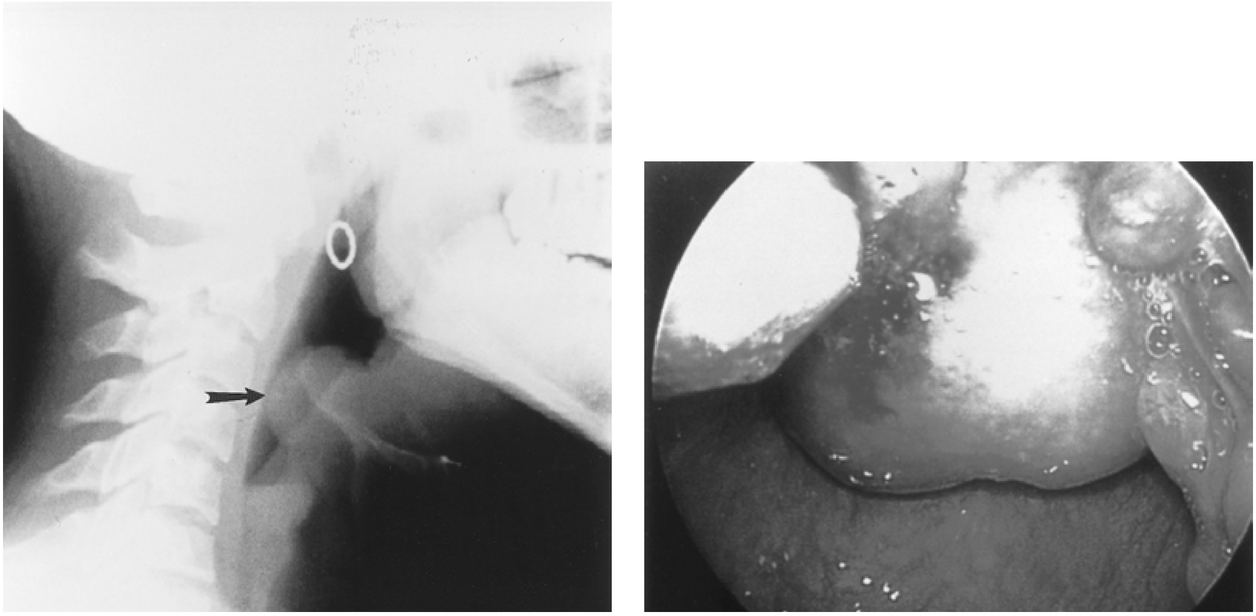
Laryngomalacia

Laryngomalacia is the most common cause of congenital stridor and accounts for ~60% of all laryngeal causes of airway obstructions. Stridor typically begins after birth and worsens during the first several months of life. Classically, the stridor is inspiratory, worse in the supine position and with crying, and better with prone positioning. The pathophysiology of laryngomalacia is most likely multifactorial and includes redundant laryngeal soft tissue, poor cartilaginous support, inadequate neurological control, and a narrow omega-shaped epiglottis. An association with GERD has also been proposed.

The diagnosis of laryngomalacia is best made by FFL in an awake child. Common endoscopic findings include collapse of the supra-arytenoid tissue, arytenoepiglottic folds, and the epiglottis causing obstruction of the airway (**Fig. 18–1**). As previously discussed, strong consideration should be made for a formal DLB in all but the most routine and mild cases to rule out a concomitant airway lesion (10%). Most children (90%) require no therapy for isolated laryngomalacia and will have spontaneous resolution usually before age 1 year. Up to 10% will have signs of severe airway obstruction or failure to thrive and require surgical therapy. Supraglottoplasty, the most commonly performed procedure for this condition, is the removal of the offending redundant supraglottic tissue, which can vary between patients. This procedure can be performed either with a CO_2 laser or with microlaryngeal instruments.

Acute Supraglottitis (Epiglottis)

Acute supraglottitis is a rare bacterial infection typically caused by *Haemophilus influenzae* type B (HIB). Other pathogens, such as *Staphylococcus* and *Haemophilus parainfluenzae,* have also been implicated in supraglottitis, particularly in older patients. The incidence of this potentially life-threatening disorder has dramatically decreased since 1987, when the first conjugate HIB vaccines were introduced. Today, most otolaryngology residents will never see a young child with supraglottitis in their training.



A B

Figure 18–2 *Lateral soft tissue radiographs of the neck reveal the classic "thumbprint" sign of the swollen epiglottis, indicative of acute supraglottitis.*

Children with supraglottitis usually present between the ages of 2 and 6 years with a history of a rapidly progressive upper respiratory infection, with high fevers (102–103°F), severe odynophagia, restlessness, drooling, and a muffled voice. The patient will often be in the upright position, with the head held forward and the tongue protruding from the mouth. Marked inspiratory stridor, retractions, and tachypnea are common. Physical examination should be limited to an overall inspection and auscultation of the heart and lungs. The child should not be traumatized, particularly with a vigorous intraoral examination or by intravenous puncture. The only diagnostic test that might be appropriate is a lateral soft tissue x-ray of the neck to evaluate the epiglottic shadow (**Fig. 18–2**).

The child should be taken immediately to the operating room to undergo slow induction inhalation anesthesia while in a parent's arms. Once a mask airway is obtained, intravenous line placement and blood cultures are obtained. Intraoral endotracheal intubation is performed by the most experienced physician, after which epiglottic cultures can be obtained. Blood and epiglottic cultures and sensitivity tests will usually confirm the diagnosis and guide antibiotic therapy. Laryngoscopes, rigid ventilating bronchoscopes, and tracheotomy equipment should be available immediately in the operating room. Patients treated in the intensive care unit with intravenous antibiotics typically respond within 48 hours and can be extubated.

Glottic/Subglottic

Laryngotracheobronchitis (Croup)

Acute laryngotracheobronchitis (LTB) is a common cause of stridor in children, usually occurring prior to age 2 years. LTB represents an acute viral illness usually caused by parainfluenza virus and presents with diffuse edema of the lower airway. The symptoms occur gradually over several days and may be associated with signs of an upper respiratory tract infection. Children demonstrate a "barking" cough with biphasic stridor due to subglottic edema. Few children develop airway obstruction severe enough to require endotracheal intubation. Most patients are less than 2 years old, although older children may be susceptible due to congenital narrowing of the subglottis (i.e., elliptical malformation).

Anteroposterior radiographs of the neck often reveal the narrowing of the subglottic tissues, termed the "steeple sign." Although usually not required, FFL reveals soft tissue swelling just below the true vocal cords (TVCs). No significant leukocytosis is noted on complete blood count. Most cases can be treated expectantly on an outpatient basis. Only ~10% of cases will require hospitalization for humidification, supplemental oxygen, fluids, racemic epinephrine, and corticosteroids, although the use of the latter is controversial. Airway support or pulmonary failure is considered uncommon.

Subglottic Stenosis

Subglottic stenosis (SGS) is the most common fixed anatomical airway obstruction. This is due to the narrow confines of this portion of the airway and the noncompliant circumferential cartilaginous support that is unique to the subglottis. The etiology is either from congenital narrowing or from soft tissue formation (acquired) after endotracheal intubation. No matter the etiology, the symptomatology in patients with SGS includes biphasic stridor, retractions, and a normal voice or cry. Respiratory distress and feeding difficulty may be present in severe cases. In milder cases, patients may be symptomatic only during bouts of upper respiratory infections. The diagnosis of SGS may be presumed from history, physical examination, flexible fiberoptic laryngoscopy, and AP radiographs of the neck, but surgical endoscopy is required to confirm the size of the airway. The stage of SGS is based on the calculated airway size reduction determined during surgical endoscopy, as previously described (**Table 18–4**).

Previously, cases of acquired SGS were thought to outnumber congenital SGS by 9:1, although congenital SGS may be more prevalent than once thought. Since the widespread use of endotracheal intubation in 1965, the number of cases of acquired SGS dramatically increased but has subsequently stabilized due to the improvements in intensive care. Damage from the endotracheal tube causes mucosal edema, ulceration, granulation tissue, and subsequent scar formation (**Fig. 18–3**). Occasionally, minor trauma may produce subglottic cysts, which usually respond to simple laser excision. Acquired SGS is usually treated with a tracheotomy for initial airway support, with subsequent laryngotracheal reconstruction performed after 1 year of age. Long-term stenting is used based on the degree of stenosis. Cricotracheal resection has been employed recently to improve the success rate of surgical repair in cases of grade III or IV stenosis.

About 1% of "normal" children have a larynx two endotracheal tube sizes (1.0 mm) below predicted; in 0.06% of cases, the larynx is three tube sizes (1.5 mm)

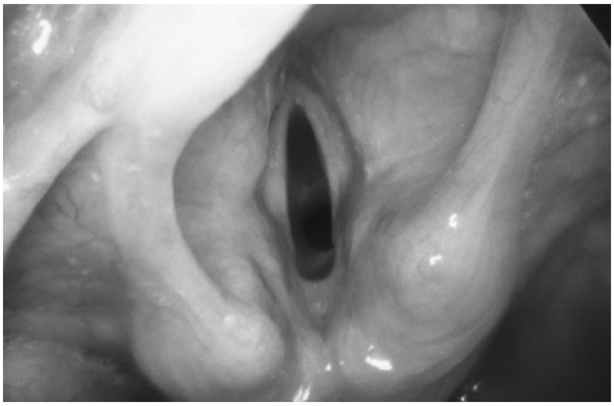


Figure 18–3 *Subglottic stenosis demonstrated on bronchoscopy after prolonged and multiple intubations.*

smaller than normal. These children are at higher risk to develop damage from endotracheal tube intubation. Congenital SGS is typically cartilaginous in nature. Anomalies consist of a thickened cricoid, trapped first tracheal ring, elliptical or ovoid shape, and, rarely, soft tissue stenosis. Patients with congenital SGS typically present at several months of age, often during an upper respiratory infection. Most patients will require only observation and symptomatic treatment for episodes of croup. Less than 10% of children will require surgical treatment, which consists of laryngotracheal reconstruction with or without tracheotomy.

Table 18–4 Stages of subglottic stenosis

Stage	Airway Reduction (%)
I	0–50
II	51–70
III	71–99
IV	100

(Data from Myer CM, O'Conner DM, Cotton RT. Proposed grading system for subglottic stenosis based on endotracheal tube sizes. Ann Otol Rhinol Laryngol 1994;103:319–323.)

True Vocal Cord Paralysis

TVC paralysis represents ~10% of stridor in infants. Congenital disorders represent over 50% of all cases of paralysis. Neurological disorders, including Arnold-Chiari malformations and Möbius' syndrome, are relatively common causes of congenital paralysis in children and may exhibit bilateral paralysis. Other patients with neurological disorders may display paradoxical motion (vocal fold dyskinesia), which is inward motion of the TVCs during inspiration. A family history may discover familiar abductor paralysis. Acquired paralysis is secondary to birth trauma, intubation, and surgical trauma, including repairs of a congenital heart defect, tracheoesophageal fistula, and most commonly a patent ductus arteriosus. As compared with adults, bilateral TVC paralysis in children is relatively more common than unilateral paralysis, although the latter is still fortunately more common. Bilateral paralysis typically produces severe airway distress and stridor due to the location of the cords in the paramedian position, while the voice (cry) and feeding are often normal. The stridor is constant, worse with exertion, and high pitched.

After complete history and physical examination to rule out concomitant anomalies, the diagnosis is confirmed by flexible fiberoptic laryngoscopy in the awake child. It is highly recommended that complete videorecording capability be available to adequately evaluate the TVC motion. In search of the underlying etiology, the clinician must investigate the entire path of the recurrent laryngeal nerve, including the brainstem, neck, and chest. Consideration should be made for MRI or CT evaluation from the brain to the chest in cases where no etiology can be determined. Surgical endoscopy should be strongly considered to evaluate the remaining airway, rule out cricoarytenoid joint fixation, and perform electromyography (EMG) of the thyroarytenoid and posterior cricoarytenoid (PCA) muscles. Possible EMG patterns include polyphasic action potentials (reinnervation), fibrillation potentials (deinnervation), and no potentials (never innervated or long-standing paralysis). Spontaneous recovery of TVC paralysis is more likely in unilateral cases with evidence of polyphasic action potentials.

Tracheotomy is often required in children with bilateral TVC paralysis, and no therapy is required for most patients with unilateral paralysis. Surgical interventions for bilateral paralysis are aimed at opening the airway by either removing or lateralizing existing tissue. Definitive treatment includes arytenoidectomy (external via laryngofissure or endoscopic laser excision), endoscopic laser cordotomy, arytenoidopexy (lateralization), and PCA reinnervation. Many patients with unilateral paralysis require no therapy or can be decannulated by age 1 to 2 years if tracheotomy is present. Definitive treatment to improve the voice and prevent aspiration in patients with unilateral paralysis is attained by medialization of the TVC externally (thyroplasty), internally (injection), or by selective reinnervation.

Recurrent Respiratory Papillomatosis

The most common tumor of the larynx, squamous papilloma, accounts for more than 80% of laryngeal growths. Recurrent respiratory papillomatosis (RRP) represents a benign tumor that recurs frequently after surgical excision and can cause complete airway obstruction. Six thousand children are affected per year in the United States, usually presenting prior to the age of 7 years, often before 2 years of age. Two broad classifications exist in patients with RRP. Juvenile, or aggressive-type, RRP has an unrelenting course of recurrent disease that may take many years to resolve and typically occurs in younger patients. Although spontaneous regression can occur around puberty, lifelong disease can be present, sometimes complicated by distal airway papillomas. Adult, or benign-type, RRP tends to have much fewer recurrences after surgical therapy, does not involve the distal airways, and tends to occur postpuberty. Because multiple surgical excisions are required during the prolonged course of many patients with this disease, over $100 million is spent annually in the United States for the management of RRP. The human papilloma virus (primarily subtypes 6 and 11) is the causative organism. An association with vaginal papillomas has been made, although the practice of elective cesarean sections has not been shown to be completely protective for the newborn.

Patients with RRP present with a hoarse or muffled voice (or cry), and some patients may be aphonic. Airway distress may be present if the papillomas obstruct the larynx. Diagnosis is made either by FFL or at the time of DLB in the operating room. Biopsy is often recommended to confirm the papilloma subtype and to rule out malignant degeneration, although the latter is a rare phenomenon. CO_2 laser laryngoscopic excision is the standard therapy. Several adjuvant therapies have been developed in light of the generally high recurrence rate, with surgery alone for those patients with "aggressive-type" disease. Interferon-α-2a has been the most widely studied adjuvant therapy, followed by indole-3-carbinol, acyclovir, ribavirin, methotrexate, retinoic acid, and most recently photodynamic therapy. All have shown initial promise, but further study is being done to confirm the role of these adjuvant therapies in the treatment of RRP. Recently another antiviral, cidofovir, has shown clinical promise in lessening the extent of disease and even eradicating papillomas.

Hemangioma

Hemangiomas are the most common tumor of the head and neck in children and represent a benign neoplastic collection of endothelial cells that undergo an initial rapid growth phase during the first few years of life. Subsequently, the tumor undergoes spontaneous regression with fibrofatty tissue deposition. The classic growth pattern and characteristic appearance are adequate for the diagnosis. A biopsy of the lesion is not recommended. Most hemangiomas in children are innocuous pink or red macules that cause no dysfunction. It is estimated that 10% of the Caucasian population and as much as 22% of preterm infants will develop hemangiomas. There is a 3:1 female preponderance as well as a predilection for Caucasians.

When the lesions are cutaneous, the diagnosis is easily made. Skin lesions can vary from small red macules to massive tumors that distort normal features

Figure 18–4 *Subglottic narrowing due to a subglottic hemangioma. Note the erythematous soft tissue stenosis and narrowing of the airway.*

and cause compression of the visual axis or aerodigestive tract. Multifocal local and systemic disease that requires active therapy occurs in less than 5% of all cases. Stridor can be caused by subglottic hemangiomas, which is associated with cutaneous hemangiomas in 50% of children (**Fig. 18–4**). These children present with airway symptoms similar to patients with subglottic stenosis. A severe and potentially life-threatening complication that may occur in a hemangioma-like lesion (hemangioendothelioma) is Kasabach-Merritt syndrome. This is a consumptive coagulopathy that occurs within these large vascular tumors.

Treatment depends on the severity of the lesions. Careful observation in patients with small macular lesions is still recommended. At the first sign of rapid growth either medical or laser therapy should be instituted. Some patients present with massive tumors or airway lesions already present. Medical therapy has traditionally consisted of high dose (2–4 mg/kg/day) corticosteroids given for weeks to months based on the response to therapy. Complete response to corticosteroid therapy is reported in only about one third of cases. Recently, interferon-α-2a was reported to induce complete responses in up to 60% of patients. Due to the possibility of neurological complications (fine motor and gait dysfunction), particularly in patients under 1 year of age, dosages should be gradually increased up to a goal of 3 million units/m^2. The duration of interferon therapy in massive hemangiomas can be for up to 16 months. Laser therapy, such as with the yttrium-aluminum-garnet (YAG) laser, flash lamp pumped dye, and copper vapor lasers, have been used successfully to treat superficial hemangiomas, but they are limited by the depth of penetration of the laser. Therefore, early recognition and treatment are vitally important for laser therapy to be successful. Recently, intralesional YAG laser therapy has been proposed in massive tumors. This treatment holds promise, but further studies are required.

Tracheal

Tracheomalacia/Tracheal Stenosis

Obstruction of the trachea by either functional collapse (malacia) or a fixed narrowing (stenosis) is generally considered a rare cause of significant airway compromise in children. Tracheomalacia may be primary or secondary to external compression. Primary tracheomalacia may be related to cartilaginous weakness that has been hypothesized to be a factor in the pathogenesis of laryngomalacia. The narrowing may be generalized or segmental. External compression is most commonly caused by anomalous vascular structures, or more rarely by mediastinal tumors or bronchogenic cysts. Common vascular anomalies include aberrant innominate artery, left pulmonary artery, left subclavian artery, and aortic arch and cardiac anomalies. Stenosis represents a congenital narrowing of the trachea either by scar tissue or due to complete tracheal rings. The narrowing may be segmental, funnel-like, or generalized in nature.

Stridor is primarily expiratory and may be associated with a deep seal-like cough. Patients may experience cyanosis, apneic spells, or recurrent lower respiratory infections. The diagnosis is confirmed by surgical endoscopy, with radiographic imaging (MRI or CT) reserved for patients with external tracheobronchial compression. Primary tracheomalacia usually resolves spontaneously during the first year of life. Patients with severe primary tracheomalacia require continuous positive airway pressure via nasal catheter or tracheotomy. Surgical intervention is typically required in patients with vascular compression or tracheal stenosis. Repair of the former consists of either aortopexy (innominate artery compression) or complex cardiac reconstruction. Segmental or funnel-type tracheal stenosis, fortunately, is the most common and requires a tracheoplasty using either cartilage interposition grafts or pericardial patches. In addition, resection with primary anastomosis has been successfully used for short-segment tracheal stenosis. Tracheal homograft transplant has also been advanced in Europe as a successful treatment, particularly in long-segment stenosis.

Figure 18–5 *Lateral chest radiograph of a 2-year-old child with an upper esophageal foreign body (coin). Note the secondary narrowing of the trachea due to edema of the tracheoesophageal membrane.*

Foreign Bodies

Inhaled or ingested foreign objects are a relatively common cause of airway obstruction and can be associated with significant mortality and morbidity. Public awareness and prevention education, rapid-response paramedic teams, and the introduction of the Heimlich maneuver have all decreased the mortality of children with foreign bodies in the United States from 650 deaths in 1968 to 261 deaths in 1990. Although a variety of foreign objects have been aspirated or ingested, esophageal coins are most commonly found. Airway obstruction occurs due to secondary swelling of the posterior tracheal wall (**Fig. 18–5**). Due to either the immediate demise of the patient or the rapid expulsion of the object, laryngopharyngeal foreign bodies rarely present to the clinician. Tracheobronchial foreign bodies primarily consist of vegetative matter, primarily peanuts, followed by a variety of inanimate objects (e.g., beads, toys).

Esophageal foreign bodies may be asymptomatic for long periods of time, until airway distress, dysphagia, and drooling develop. Pulmonary objects often produce a period of violent paroxysms of coughing, choking, and occasionally a color change. This acute episode may go unnoticed and is followed by an asymptomatic stage that may last for weeks to months. The final stage is that of intrathoracic complications, such as airway obstruction, infection, or tracheal erosion. Chest radiographs, with inspiratory and expiratory views, will often demonstrate a unilateral infiltrate or air trapping. Because the right mainstem bronchus take-off is at less of an angle to the trachea, most foreign bodies tend to lodge in the right bronchus intermedius or segmental bronchi. A high index of clinical suspicion should be maintained for airway foreign bodies in any child with atypical, unilateral, or suspicious respiratory symptoms, so as to initiate prompt treatment and avoid possibly devastating complications.

The treatment of aerodigestive foreign bodies involves prompt endoscopic evaluation and removal of the object. For nontraditional objects, the surgeon should obtain a duplicate object and practice the extraction technique. Endoscopic forceps offer state-of-the-art visualization and extraction capabilities. Fogerty catheters are often helpful to either manipulate the foreign body to a more proximal location or stabilize the object for endoscopic removal. Vegetative matter is especially difficult to remove because of the soft and immunoreactive nature of the fragments. Bronchoalveolar lavage is often required to assist in clearing the debris from the airway. Other adjuvant therapies include antibiotics, mucolytics, bronchodilators, and chest physiotherapy. Although rare, the protocol for laryngeal foreign bodies should be similar to that for the careful and gentle approach to patients with epiglottitis. The otolaryngologist, with full airway support equipment available, should perform laryngeal foreign body extraction.

Suggested Readings

Albert D, Leighton S. Stridor and airway management. In: Bailey BJ, ed. Otolaryngology–Head and Neck Surgery. 2nd ed. Philadelphia: Lippincott-Raven; 1998

Cotton RT, Reilly JS. Stridor and airway obstruction. In: Bluestone CD, Stool SE, Kenna MA, eds. Pediatric Otolaryngology. 3rd ed. Philadelphia: WB Saunders; 1996

Greinwald JH, Smith RJH. Laryngeal disorders. In: Oski FA, DeAngelis CD, Feigin RD, McMillan JA, Warshaw JB, eds. Principles and Practice of Pediatrics. Philadelphia: Lippincott; 1999

Lesperance MM, Zalzal GH. Assessment and management of laryngotracheal stenosis. Pediatr Clin N Am 1996;43: 1413–1427

Pransky SM, Albright JT, Magit AE. Long-term follow-up of pediatric recurrent respiratory papillomatosis managed with intralesional cidofovir. Laryngoscope 2003;113:1583–1587

Roger G, Denoyelle F, Triglia JM, Garabedian EN. Severe laryngomalacia: surgical indications and results in 115 patients. Laryngoscope 1995;105:1111–1117

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Which of the following statements is correct?
   A. Stertor is a high-pitched musical noise.
   B. Rhonchi are high-pitched noises caused by reverberation of oropharyngeal tisues.
   C. Rales are a popping noise heard in the peripheral lung fields.
   D. Stridor, wheezing, and rales all refer to the same pathological airway noises
2. Which of the following types of radiographic evaluations may be used to evaluate patients with airway noise?
   A. Barium swallow study
   B. CT
   C. MRI
   D. Ultrasound
   E. All of the above
3. Which of the following is *not* a part of the CHARGE association?
   A. Choanal atresia
   B. Colobomas of the eyes
   C. Cardiac defects
   D. Hyalinization of the glottic tissue
   E. Ear malformations

Chapter 19

CLINICAL GENETICS IN OTOLARYNGOLOGY

Simon I. Angeli, Nancy Sculerati, and Thomas R. Van De Water

Genetics is playing an increasingly critical role in the practice of clinical medicine. This is due in part to the importance that knowledge of genetics has in the treatment of genetic diseases, but also in part to the fact that such knowledge provides an understanding of the fundamental biological process of most diseases. Some common disorders are due to the interaction of multiple genes and environmental factors, and genetic variations may have either a protective or a pathological role in the expression of a disease. For example, although presbycusis and noise-induced hearing loss could be dismissed as the result of accumulated age and environmental trauma, animal studies have shown that genetic background is an important determinant of the final outcome. Today's otolaryngologist must understand the science of medical genetics because this knowledge can lead to more effective disease diagnosis, treatment, and prevention. Hearing loss, neoplastic disease, metabolic disease, coagulopathies, and congenital malformations of the head and neck are the disorders in otolaryngology most likely to involve genetic variations, and in these instances knowledge of genetics and molecular biology may be fundamental for making a differential diagnosis and selecting a management protocol.

HUMAN MEDICAL GENETICS

Genetic information is passed from one generation to the next as coded information in the human genome. Deoxyribonucleic acid (DNA) sequences form genes that are expressed through ribonucleic acid (RNA) into proteins. The human genome includes all DNA in the 46 chromosomes: 22 pairs of autosomes and the sex chromosomes XY in males and XX in females. An estimated 30,000 to 35,000 genes are present on the human genome, but they account for only ~2% of its length (see Chapter 15). It is currently accepted that more than one protein can derive from one gene through the mechanism of alternative splicing. Scattered within and between genes are vast stretches of DNA whose function is less well understood.

Genes are distributed unevenly between chromosomes and within each chromosome. Chromosomes 17, 19, and 22 are particularly gene dense as compared with other chromosomes, such as 4, 8, 13, 18, and Y. Within each chromosome, gene density is highest in areas rich in the DNA bases cytosine and guanine. Gene locus refers to the position of the gene in a chromosome. An alphanumeric code describes the locus of a gene on each nuclear chromosome. The first number (or letter) designates the chromosome: autosomes 1 through 22, sex chromosomes X or Y. The next part of a gene locus is a letter: *q* for the long arm of a chromosome, *p* for the short arm. The final part of the locus code is a number that often contains a decimal point. This refers to a Giemsa-stained band on the arm of the designated chromosome. Several genes that are involved in energy metabolism are located not on chromosomes located within the nucleus but rather on a mitochondrial chromosome; these are designated mitochondrial DNA (mtDNA).

Individuals inherit 50% of their chromosomal makeup from their father and 50% from their mother. Every gene is paired, one copy of the pair being paternal and the other maternal in origin. The two genes of the gene pair are not necessarily identical; subtle differences may be present. Each copy of a gene is known as an allele. Certain alleles are associated with a change in gene function that is reflected by a specific physical characteristic or phenotypic appearance. Disease-causing alleles arise from mutations. Mutant genes can occur in the germ line (hereditary) or can arise de novo from changes in DNA sequences (sporadic). Different mutations in the same gene can act as alleles and cause different manifestations of disease (e.g., familial medullary carcinoma and multiple endocrine neoplasias). Current data indicate that mutations known to cause diseases have been identified in nearly 1000 genes.

Genetic disorders are broadly classified into four major groups:

- Chromosome disorders, in which entire chromosomes or large segments of them are altered. These changes are often visible by karyotyping, and may be inherited in a Mendelian fashion or may represent isolated genetic mutations. Examples of chromosome disorders include Down syndrome and Turner syndrome.

- Monogenic disorders, also known as Mendelian disorders, because they involve alterations in single nuclear genes and follow Mendel's laws of inheritance. Examples include nonsyndromic sensorineural hearing loss and cystic fibrosis.
- Multifactorial disorders, which are due to a combination of multiple genetic as well as environmental causes. Well-known examples include birth defects such as cleft lip and palate, as well as some adult disorders such as heart disease, diabetes, and cancer.
- Mitochondrial disorders, which are single-gene disorders caused by alterations in the mitochondrial chromosome. Each mitochondrion contains two to 10 mitochondrial chromosomes. Human mtDNA encodes 13 messenger RNA (mRNA) molecules, as well as 2 ribosomal RNAs (rRNAs) and 22 organelle-specific transfer RNAs (tRNAs), which are required for assembling a functional protein-synthesizing system. MtDNA, along with its variations, is inherited solely from the mother because ova are rich in mitochondria and sperm are not. Therefore, it follows a maternal pattern of inheritance, which is different from the pattern of inheritance of nuclear genes (**Fig. 19–1**). Examples include familial susceptibility to aminoglycoside ototoxicity and presbycusis.

Mendelian disorders traditionally have received the greatest amount of attention since the establishment of the interaction between alleles and phenotypic appearance by Gregor Mendel in the 19th century. Alleles that express in only a single gene (i.e., heterozygous) are dominant inheritance. Recessive alleles are only expressed if present doubly (homozygous). These patterns of allelic inheritance are referred to as autosomal dominant and autosomal recessive, respectively. Autosomal dominant inheritance is characterized by transmission of the disease phenotype generation after generation, with an equal number of affected males and females and a recurrence risk of 50%. Autosomal recessive inheritance is characterized by "skipped" generations (the disease is seen in multiple siblings, but usually no earlier generations are affected), no sex predilection, and a recurrence risk of 25%. Consanguinity is usually present in recessive disease because related individuals are more likely to share the disease gene.

Fewer genes exist on the Y chromosome than on the X chromosome; therefore, sex chromosome alleles usually are not paired. X-linked patterns of inheritance usually are recessive. This implies that a female, by virtue of two X chromosomes, is a carrier. Fifty percent of a carrier female's sons are affected, and 50% of her daughters are

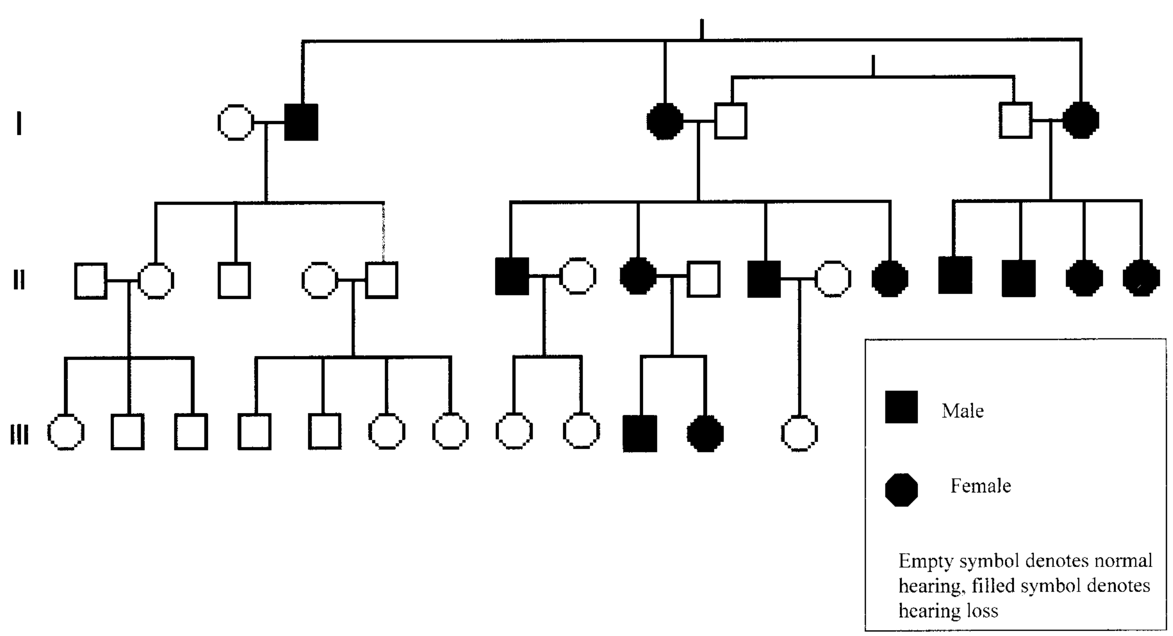


Figure 19–1 *A family tree depicting the inheritance of deafness for three generations of a family carrying a mutation in their mitochondrial DNA. Note that both male (squares) and female (circles) family members are affected but that the pattern of inheritance is strictly maternal. This is because only the fertilized ovum contains most of the heritable mitochondrial DNA, and the contribution of mitochondrial DNA to the fertilized zygote from the spermatazoa is negligible.*

carriers. All male progeny of affected males are normal, and all female progeny of affected males are carriers.

Genetic diseases usually show variable expressivity beyond simple Mendelian inheritance. An affected individual may exhibit few, some, or all of the manifestations of an allele. Occasionally, an individual with a particular gene abnormality will not exhibit the disease phenotype at all, even though the person can transmit the disease gene to the next generation, and the gene is said to have reduced penetrance. Variable expression of a genetic disease may be caused by genetic or environmental factors. Among the genetic factors, modifier genes, allelic heterogeneity, and genomic imprinting are common. Other genes can influence the expression of a disease-causing gene and are termed modifier genes. Allelic heterogeneity refers to the effect that different types of mutations within the same gene (i.e., different alleles) can have on the phenotypic expression. An example of allelic heterogeneity is sensorineural hearing loss due to *MYO7A* gene defects: separate mutations in this gene are responsible for autosomal recessive nonsyndromic congenital deafness, autosomal dominant nonsyndromic progressive deafness, and syndromic deafness with blindness (Usher syndrome type 1B). Genetic imprinting is another form of phenotypic variation. An imprintable allele will be transmitted in a Mendelian mode, but expression will be determined by the sex of the transmitting parent. Paternal imprinting is used to imply that there will be no phenotypic expression if the disease allele is transmitted from the father; his offspring, however, will be nonmanifesting carriers. Other sources of modification of the effect of a gene are alternative splicing, uniparental disomy (both members of an allele pair derive from one parent), and "epigenetic" phenomena such as methylation and histone modification. Additionally, the expression of a gene may depend on an environmental factor; for example, a tumor from a proto-oncogene after carcinogen exposure and susceptibility to aminoglycoside ototoxicity due to an mtDNA mutation.

The most comprehensive reference source for genetic diseases for clinicians, *Mendelian Inheritance in Man,* edited by Victor A. McKusick, was reviewed for the creation of this chapter and is available in published book form in medical libraries and online at http://www.ncbi.nlm.nih.gov/Omim/, a Web site supported by the National Institutes of Health.

TUMORS OF THE HEAD AND NECK

Cells develop, differentiate, grow, and die in response to a complex system of biochemical signals. Neoplastic cells usually contain mutations that free them from the normal constraints thereby allowing them to proliferate uncontrollably. Although virtually all neoplasms are genetic, the overwhelming majority of mutations in tumors are found only in a patient's neoplastic cells and are not inherited. Mutations in somatic cells are caused by intrinsic errors in DNA replication or repair, or are induced by carcinogen exposure. However, ~1% of all cancers arise in patients with germ line mutations that predispose them to cancer, but even in this situation additional somatic cell mutations are required for a tumor to develop. These patients have a hereditary cancer syndrome.

Three classes of genes are involved in both sporadic and hereditary neoplasia: (1) those that normally inhibit cell proliferation (tumor suppressor genes), (2) those that activate cell proliferation (oncogenes), and (3) those that participate in DNA repair.

- Tumor suppressor genes stop tumor growth. Loss-of-function mutations that deactivate or downregulate tumor suppressor genes are associated with neoplasia. These mutations are usually recessive; both normal gene copies must be lost for disease to develop. The gene for p53 is an example of a tumor suppressor gene.
- Oncogenes initiate and promote neoplastic cell growth. Mutations that activate or amplify oncogenes are associated with tumors. Most oncogenes originate from proto-oncogenes, which are genes involved in the basic regulation processes of cell growth. When a mutation occurs in a proto-oncogene, it can become an oncogene. *RET* is an example of a dominant oncogene whose deregulated expression drives cell transformation. Multiple mutations are seen in most malignancies. Inactivation of tumor suppressor genes coupled with activation of oncogenes produces malignant neoplasia.
- Inherited defects in DNA repair during DNA replication can lead to a high frequency of somatic mutations. Familial breast cancer is an example of a somatic mutation affecting DNA repair in cell regulatory pathways.

Features of Hereditary Neoplasms (Mutations in Patient's Germ Line)

Hereditary tumors tend to have a different clinical picture than sporadic tumors of the same histology. Familial tumors typically exhibit some or all of the following clinical characteristics:

- Bilateral or multifocal sites
- Presentation at a relatively young age

- Aggressive behavior (large size and/or late stage at presentation with increased chance of dissemination and recurrence)

Medullary Thyroid Carcinoma and Parathyroid Tumors

Gene map locus: 10q11.2

Gene: *RET*

Autosomal dominant

About 25% of medullary thyroid carcinoma (MTC) tumors are multifocal. These hereditary tumors occur in patients with germ line mutations in the *RET* gene. The *RET* gene encodes a receptor tyrosine kinase that is involved in embryonic neural crest cell migration. Gain-of-function mutations in *RET* result in excess tyrosine kinase activity and increased signal transduction, leading ultimately to cellular proliferation. Elevated calcitonin levels and genetic testing for specific *RET* gene mutations can be used to identify affected family members. Gain-of-function mutations can produce any of three autosomal dominant inherited cancer syndromes, depending on the type and location of the mutation:

- MTC alone (without any other associated anomalies) in familial MTC
- Multiple endocrine neoplasia (MEN) type 2A: MTC, pheochromocytoma, and hyperparathyroidism (parathyroid hyperplasia)
- MEN2B: MTC and pheochromocytoma occur, but these patients have a characteristic syndromic appearance with a marfanoid habitus and mucosal ganglioneuromas of the lips, tongue, and intestinal tract. Parathyroid hyperplasia is not a feature. MEN2B is the most aggressive of the three familial forms of MTC; metastases occur in children by the age of 5 years

Genetic screening for RET gene mutations in family members of any of the three MTC syndromes is available. Positive results are an indication for prophylactic thyroidectomy.

Familial Papillary Carcinoma of the Thyroid

Gene map locus: l0ql l–q12

Although nonmedullary thyroid cancer rarely has a hereditary basis, nonmedullary thyroid cancers do occur in hereditary cancer syndromes such as familial polyposis coli.

An autosomal dominant familial papillary carcinoma has been reported. These families have multiple members with a history of papillary carcinoma of the thyroid, usually aggressive. Inheritance appears to be via a dominant oncogene.

Head and Neck Squamous Cell Carcinoma

Tumor protein p53, TP53

Gene map locus: l7p3l.1

Squamous cell carcinoma cells contain abnormal DNA. Prognosis of head and neck cancer patients can be estimated in some cases by analysis of serum and plasma DNA released into the circulation, showing DNA mutations that indicate neoplastic progression. The presence of aggressive head and neck squamous cell carcinomas in young adults (particularly in those patients without lifestyle risk factors) suggests genetic predisposition.

The normal gene encoding p53 protein acts as a tumor suppressor gene; p53 is a nuclear protein that arrests the cell cycle at Gl in response to damaged DNA. Alterations of the *p53* gene or its encoded protein are the most common genetic abnormalities observed in human cancers.

- Approximately 50% of head and neck squamous carcinomas contain *p53* mutations.
- The *p53* gene is likely a molecular target of tobacco and alcohol, which have been associated with head and neck squamous cell carcinoma; *p53* mutations are much more common among users of both tobacco and alcohol.
- In patients with advanced head and neck cancer, at least 50% of cases with complete resection on the basis of negative histopathological margins have at least one surgical margin positive for the same *p53* mutation found in the tumor, correlating with recurrence.
- Germ line *p53* mutations predispose primarily to osteosarcoma, soft tissue sarcoma, brain tumors, leukemia, and breast cancer in women.
- In sporadic tumors, tumor histology often varies with both *p53* sequence mutation and environmental risk factor. The specific mutations in *p53* vary in lung cancers associated with cigarette smoke exposure, hepatoma associated with aflatoxin exposure, and skin cancer associated with ultraviolet (UV) radiation. The *p53* mutations induced by UV radiation are also present in actinic keratosis. In squamous cell carcinoma of the skin, sunlight can act as both a tumor initiator and a tumor promoter.

Nasopharyngeal Carcinoma

A high frequency of nasopharyngeal carcinoma (NPC) in ethnic Cantonese people and the existence of families with multiple members with NPC suggest genetic factors. The frequency of NPC is almost 100 times greater in southern Chinese people than in European populations. Males are more often affected than females. The overall age group affected is younger than most adult cancer patients.

Epstein-Barr virus (EBV) infection may act as an initiator of neoplasia in genetically susceptible individuals. EBV infection is an early event in NPC. All NPC tumor cells contain EBV DNA. Preinvasive lesions of the nasopharynx (dysplasia, carcinoma in situ) are also infected with clonal EBV.

Lymphoma

Lymphoma patients often present to the otolaryngologist, either because of accessible cervical adenopathy for tissue diagnosis or with a primary lymphoma of the salivary gland, tonsil, or sinus. Non-Hodgkin's B-cell lymphomas are most common in the head and neck.

Non-Hodgkin's B-cell lymphomas frequently contain translocation mutations that activate oncogenes. Mutations in follicular non-Hodgkin's B-cell lymphomas involve activation of the mitochondrial *BCL2* gene, reducing programmed cell death (apoptosis).

Many lymphomas of the head and neck show genetic mutations in association with EBV infection. Burkitt's lymphoma is characterized by translocations that activate the *MYC* proto-oncogene on chromosome 8 and by EBV infection.

Alteration of the tumor suppressor gene *p53* is frequent in acquired immunodeficiency syndrome–related non-Hodgkin's lymphomas. EBV sequences are detected in 40 to 60% of acquired immunodeficiency syndrome (AIDS)-related non-Hodgkin's lymphoma.

Rhabdomyosarcoma

Rhabdomyosarcoma (RMS) is the most common sarcoma of childhood, and its most common site of presentation is the head and neck region (40% of cases). The two major subtypes of pediatric RMSs are embryonal and alveolar. The alveolar subtype has a worse prognosis.

In alveolar RMS, fusion of a homeobox gene (*PAX3* or *PAX7*) with one of the forkhead transcription factor genes occurs in tumor cells. Molecular biology assays on frozen sections of surgical specimens can increase the accuracy of histological diagnosis in sarcomas, which can be difficult to distinguish on histological appearance alone.

Hereditary Paraganglioma

Gene map locus: 11q23

Gene: *SDHD*

Autosomal dominant

The *SDHD* gene encodes the small subunit of cytochrome b (cybS) in succinate-ubiquinone oxidoreductase in the mitochondria. In families with hereditary paraganglioma, germ line mutations in the *SDHD* gene have been reported. Mitochondria may play an important role in the pathogenesis of certain tumors, and cybS plays a role in normal carotid body physiology. Germ line loss of function mutations in the paternal alleles and subsequent somatic loss of normal maternal alleles suggest that *SDHD* functions as a tumor suppressor gene at the cellular level and needs two events for inactivation. It has been suggested that on the basis of the phenotypic similarity between paraganglioma tumors and the normal carotid body exposed to chronic hypoxia, cybS is a critical component of the oxygen-sensing system of paraganglionic tissue, and that its loss may lead to chronic hypoxic stimulation and cellular proliferation.

Carotid body tumors and glomus jugulare tumors are called paragangliomas because they arise from paraganglionic tissue. They are also called chemodectomas (because paraganglia are chemoreceptor structures). Clinical features of hereditary paragangliomas include

- Familial carotid body tumors are often multiple and may be bilateral. Bilateral disease occurs in almost a third of known familial cases, but in less than 5% of sporadic cases.
- Familial glomus jugulare tumors are rare.
- Hereditary paragangliomas appear to be inherited almost exclusively via the paternal line, indicating genomic imprinting. Patients with hereditary paragangliomas inherit the disease gene from their father; tumors are not observed in the offspring of an affected female (until subsequent transmission of the gene through a male descendant).

Neurofibromatosis Type 2, Central-Type Neurofibromatosis, Bilateral Acoustic Schwannomas, and Acoustic Neurofibromatosis

Gene map locus: 22q12.2

Autosomal dominant

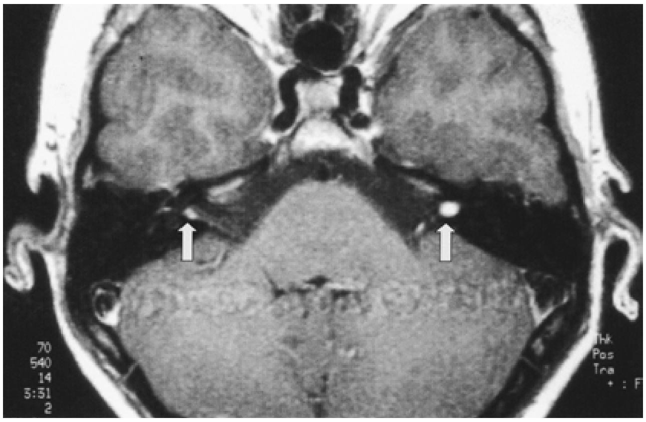


Figure 19–2 *A gadolinium-enhanced magnetic resonance image (MRI) (T1 weighted, axial plane) through the internal auditory canal (IAC) of an asymptomatic 13-year-old patient with a family history of neurofibromatosis type 2 (NF2). This screening MRI shows bilateral enhanced masses within the patient's IACs consistent with the presence of bilateral vestibular schwannomas (arrows) and confirms a diagnosis of NF2.*

About 4% of all cases of acoustic neuroma are bilateral. Bilateral tumors

- Are inherited in an autosomal dominant pattern
- Frequently have been reported to be massive at presentation
- Usually are associated with a central form of neurofibromatosis (NF2)

NF2 is characterized by additional central nervous system (CNS) tumors (meningiomas of the brain, and schwannomas of the dorsal roots of the spinal cord). NF2 differs from the peripheral form of neurofibromatosis (von Recklinghausen's disease); typically, there are no cafe-au-lait spots or peripheral neurofibromata in NF2. Gadolinium-enhanced magnetic resonance imaging and neurological and ophthalmologic evaluations are indicated in NF2 patients and their relatives (**Fig. 19–2**). The generally accepted criteria for NF2 diagnosis include bilateral vestibular schwannomas or family history of NF2 in one or more first-degree relatives plus either (1) unilateral vestibular schwannoma at age less than 30 years or (2) any two of the following: meningioma, glioma, schwannoma, or juvenile posterior subcapsular lenticular opacities/juvenile cortical cataract.

COAGULOPATHIES

Because of the vascularity of the head and neck, the otolaryngologist often encounters patients with bleeding disorders, presenting either with operative complications or with a primary problem of spontaneous bleeding (e.g., epistaxis or hemoptysis). Many coagulopathies are hereditary. Both prospective and retrospective studies of post-tonsillectomy hemorrhage have supported the cost effectiveness of restricting preoperative coagulation testing to patients with a personal or a family history of excessive bleeding. A hematology consult is invaluable in the diagnosis and management of patients with coagulopathies.

Von Willebrand Disease

Gene map locus: 12p13.3

Autosomal dominant

Von Willebrand's disease is the most common inherited bleeding disorder, with an estimated prevalence as high as 1% of the population. The severity of the coagulopathy is markedly variable both between patients and in affected individuals over time.

- Bleeding is due to abnormalities of von Willebrand factor (a plasma protein essential for both platelet adhesion and the stabilization of factor VIII).
- DDAVP® (desmopressin acetate) provides effective therapy in many patients with von Willebrand's disease, but the response is unpredictable and frequently limited: 20 to 25% of patients with the disease do not respond to DDAVP®.

Type 1 von Willebrand's disease is most common (70% of cases). There are decreased levels of von Willebrand factor. DDAVP® is most effective in the management of patients with type 1 disease. Type 2 variants (~30% of cases) generally involve a qualitative problem in von Willebrand factor activity. DDAVP® may or may not be helpful. Type 3 is rare (~1 in 1,000,000 people) and shows a complete deficiency of von Willebrand factor. Type 3 bleeding is severe, and DDAVP® is not helpful.

Selection of therapy is best chosen after identification of the specific von Willebrand disease subtype and preoperative trials of the effect of desmopressin on coagulation. Bleeding refractory to desmopressin is treated with cryoprecipitate or factor VIII. Unlike classic hemophilia, there is a prolonged bleeding time in von Willebrand disease. The prothrombin time (PT) is normal, and the activated partial thromboplastin time (PTT) may be elevated (if the factor VIII activity is 30% or less). Levels of von Willebrand factor can vary at different times in affected patients, along with bleeding time, PTT, and clinical tendency to hemorrhage.

Hemophilia A (Classic Hemophilia, Factor VIII Deficiency)

Gene map locus: Xq28

X-linked recessive

Hemophilia A is a bleeding disorder caused by a deficiency in the activity of factor VIII. The disease presents in males. Affected males have a variable phenotype of hemorrhage into joints and muscles, easy bruising, and prolonged bleeding from wounds.

There is an X-linked recessive pattern of inheritance of hemophilia A: (1) the mutant gene is located on the X chromosome, and (2) the reduction in factor VIII activity is expressed only if there is no normal allele present. There is one X chromosome in males, just one allele, and inheritance of the hemophilia gene will cause disease. In females, there are two X chromosomes, each bearing an allele. Even if the hemophilia gene is inherited, the disease will occur only if the second allele is also abnormal.

Females are therefore very rarely affected, but they are carriers. Because fathers pass a Y, not an X, chromosome to their children, hemophilia (like all other X-linked disorders) will never show male-to-male transmission.

- The severity and frequency of bleeding in hemophilia A are directly related to the amount of residual factor VIII. Replacement by intravenous factor VIII restores normal hemostasis. Factor VIII produced by the cloned gene using recombinant DNA methods avoids exposure to transfusion-associated viral diseases. Human immunodeficiency virus (HIV) infection has accounted for a significant mortality in hemophiliacs.

Venous Thrombosis Due to Factor V Leiden Deficiency

Patients with factor V Leiden gene mutations have an increased risk of venous thrombosis, particularly when other risk factors such as surgery, use of oral contraceptives, and bed rest are also present. The factor V Leiden mutation is relatively common, ranging in prevalence from 1 to 5% in the population. Diagnosis is by DNA test or by measurement of activated protein C resistance. Avoidance of risk factors and prophylactic anticoagulation therapy are recommended.

INCREASED ANESTHESIA AND OPERATIVE RISK

Malignant Hyperthermia Susceptibility

Gene map locus: 19q13.1

Autosomal dominant

Malignant hyperthermia (MH) occurs in ~1 in 50,000 general anesthesias. Elevations of creatinine phosphokinase (CPK), phosphate, and potassium in the blood indicate severe muscle damage during MH.

Children are at increased risk. A higher incidence is encountered in certain geographically defined populations (descendants of settlers of north-central Wisconsin, descendants of settlers in Quebec).

At least one form of MH is produced by a mutation in the ryanodine receptor gene (*RYR*), first isolated in families with a history of multiple relatives succumbing to MH. Inquiry about a family history for problems with general anesthesia is prudent before administration of triggering anesthetics.

Sickle Cell Anemia

Gene map locus: 11p.15.5

Gene: *HBB*

Autosomal recessive

Hereditary hemoglobinopathies affect an estimated 5% of the world's population. The hemoglobin molecule is composed of four polypeptide chains, two of which are labeled α and two labeled β. A gene in chromosome 11 encodes the β chains, and two genes in chromosome 16 encode the α chains. Sickle cell anemia (SCA) is the most common hemoglobin disorder, affecting approximately one in 500 offspring of African American births. SCA is characterized by a structural abnormality of the hemoglobin molecule due to a missense mutation that effects a substitution of valine for glutamic acid at position 6 of the β-globin polypeptide chain. The erythrocytes assume a characteristic sickle shape on conditions of low oxygen tension, leading to vaso-occlusive events, tissue infarction, pain crises, anemia, splenic dysfunction, and multiple infections. Diagnosis is done by hematologic test (hemoglobin electrophoresis) and by DNA-based detection of the hemoglobin mutation.

Hypothermia and hypoxia can precipitate sickle crisis. These risks are not present in the heterozygote (sickle trait). The homozygote requires appropriate preoperative transfusion (with consultation from hematology), attention to strict hemostasis, maintenance of normal body temperature, and avoidance of pre- or postoperative dehydration for optimal operative outcome with any surgical procedure.

Obstructive sleep apnea secondary to tonsillar and adenoidal hypertrophy is more common in children with SCA than the general pediatric population. Sensorineural hearing loss is also more common in SCA patients and may be the result of vascular insults.

INHERITED SYNDROMES ASSOCIATED WITH SINUSITIS

Cystic Fibrosis

Gene map locus: 7q3 1.2

Gene: *CFTR*

Autosomal recessive

Cystic fibrosis (CF) is the most common lethal genetic disease, affecting approximately one in 3000 Caucasian newborns, and it is less common in other ethnic groups. Decreased fluid and salt secretion occurs secondary to abnormal ion transport and causes obstructed exocrine outflow from the pancreas, excess chloride in the sweat, and the accumulation of dehydrated mucus in the airways. There is impaired exocrine function of the pancreas, intestinal glands (meconium ileus), biliary tree (biliary cirrhosis), bronchial glands (chronic bronchopulmonary infection with emphysema), and sweat glands (high sweat electrolytes, salt depletion). Pansinusitis is universal, but most often it is not accompanied by sinus complaints. Chronic *Pseudomonas aeruginosa* pneumonia is common in more severely affected patients. Sweat testing (high levels of chloride) or DNA testing can make the diagnosis in patients with mild presentations who are unrecognized and present to the otolaryngologist with chronic cough, sinusitis, and respiratory disorders. Endoscopic sinus surgery can be helpful in certain cases of CF sinusitis.

The CF gene encodes a protein labeled cystic fibrosis transmembrane regulator (CFTR). CFTR forms cyclic adenosine monophosphate (CAMP)-regulated chloride-ion channels in epithelial cells that line the bowel and respiratory tract. CFTR is also involved in regulating the transport of sodium ions. More than 700 mutations at the CF locus have been described by DNA analysis; the most common mutation (70%), labeled ΔF508, is a three-base deletion resulting in a loss of a phenylalanine at position 508 of the CFTR protein. In the homozygous form, ΔF508 results in a complete lack of chloride ion channel production, and affected individuals have a severe type of disease with pancreatic insufficiency. Other mutations result in milder phenotypes, with less severe pulmonary disease and less likelihood of pancreatic insufficiency. Carrier screening is appropriate for individuals with a positive family history of CF or for couples in which one member is a known carrier. It is common to screen for the 25 mutations that are most common in Caucasians, and this will identify 90% of CFTR mutations. There are currently clinical trials in which normal CF genes are inserted into adenovirus or other gene therapy vectors (see Chapter 15), which then may be introduced into the lungs and sinuses of CF patients.

Kartagener's Syndrome (Dextrocardia, Bronchiectasis and Sinusitis, Immotile Cilia Syndrome)

Gene map locus: 9p21–p13, 5p15–14

Autosomal recessive

- Cilia are immotile; electron microscopy of respiratory epithelium (and sperm) shows absence of dynein arms.
- Situs inversus may be present or absent, but the classic description of the syndrome includes dextrocardia, left-sided liver, right-sided stomach.
- Sinusitis and otitis media are frequent. Functional endoscopic sinus surgery has been shown empirically to be helpful in clinical reports, despite underlying lack of ciliary function in sinus epithelium. Otorrhea is common after myringotomy and tube insertion.

METABOLIC SYNDROMES INVOLVING THE HEAD AND NECK

Congenital or progressive head and neck deformities that produce upper airway obstruction, cranial nerve palsies, and hearing loss are often metabolic in-born errors. There are more than 400 known inherited enzyme deficiencies that result in lysosomal storage syndromes. Although the particular enzyme in each of these metabolic diseases is defective from birth, infants often appear normal initially. The syndromic physical characteristics become more obvious over time, as the soft tissues thicken, with cells distorted by large intracellular lysosomes. Squat, thickened facial features, short stature, hepatosplenomegaly, mental retardation, airway obstruction, and conductive hearing loss ensue.

Mucopolysaccharidosis Type I (MPS I)

Includes MPS I Hurler's syndrome, Scheie's syndrome, and Hurler/Scheie phenotype

Gene map locus: 4pl6.3

Autosomal recessive

- The autosomal recessive form of mucopolysaccharidosis (MPS) is the most frequent type of gargoylism, affecting both males and females.
- Clinical features include progressive mental retardation, growth retardation with dwarfism, upper airway obstruction, respiratory failure, hepatosplenomegaly, hearing loss.
- The enzyme deficient in Hurler's syndrome is α-L-iduronidase.

Mucopolysaccharidosis Type II (MPS II) (Hunter Syndrome)

Gene map locus: Xq28

X-linked (recessive)

- Hunter's syndrome (MPS II) affects males only.
- Is clinically less severe (and less common) than MPS I.
- Iduronate sulfatase deficiency causes tissue deposits of mucopolysaccharides and urinary excretion of large amounts of both chondroitin sulfate B and heparitin sulfate.

CHROMOSOMAL SYNDROMES

Chromosomal abnormalities can be diagnosed with a karyotype, from blood or tissue specimens properly sent to cytogenetics.

Down Syndrome, Trisomy 21

Gene map locus: 21q22.3

Down syndrome is characterized by distinctive phenotypic features caused by trisomy of all (or, less commonly, of just a critical portion) of chromosome 21. The frequency of trisomy 21 in the United States is one in 650 to one in 1000 live births. The risk of having a child with trisomy 21 increases with maternal age. Phenotypic features include

- Developmental delay and premature "senility" (neuropathologic hallmarks of Alzheimer's disease: senile plaques and neurofibrillary tangles) are usually present in the brain of individuals with Down syndrome by the age of 40 years.
- Characteristic facies: flat skull base, short nasal bridge, macroglossia
- Thick cerumen and external auditory canal stenosis
- Conductive and mild sensorineural hearing losses are more common than in the general population.
- Upper airway obstruction including sleep apnea is also more common, likely due both to the relative small size of the airways and to hypotonia.
- Leukemia has increased incidence in Down syndrome.
- Both hearing loss and upper airway obstruction can impede intellectual development and require aggressive and early treatment for optimal outcome in these patients.

Features of trisomy 21 that affect surgical planning include

- Neck extension can cause subluxation of the cervical spine due to an anomaly of C1. Careful and conservative positioning during tonsillectomy and neck surgery is required.
- Major congenital cardiac malformations (particularly the atrioventricular canal) occur frequently, and cardiac anomalies overall exist in 40% of Down patients. A careful cardiac history before any operative procedure with general anesthesia may avoid unexpected hemodynamic compromise. Results of an echocardiogram are often helpful. Antibiotic prophylaxis against endocarditis is commonly required because of atrial or ventricular septal defect, or the presence of graft material.

Turner Syndrome, X Chromosome Deficiency (XO)

This chromosomal abnormality affects females only. Approximately 50% of cases have a 45, XO karyotype; the remainder of cases have structural abnormalities of the second X chromosome or mosaic cell lines.

Turner syndrome is the most common sex-chromosome abnormality in females, affecting one in 1500 to 2500 live-born female infants.

Features are variable but always include

- Short stature. Failure to enter puberty because of gonadal dysgenesis is usual; hormone administration helps provide development of secondary sex characteristics.
- Normal intelligence

Other features may include coarctation of the aorta, neck webbing, and lymphedema.

Primary hypothyroidism develops in 10 to 30% of cases (Hashimoto's thyroiditis).

Otitis media and associated conductive hearing loss are frequent, and aggressive intervention is recommended. There is a high rate of acquired cholesteatoma. Middle ear disease resembles that of a cleft palate population.

In addition to conductive hearing loss associated with middle ear disease, there is a high incidence of progressive sensorineural hearing loss by early adulthood.

CRANIOFACIAL DYSMORPHISM SYNDROMES

Cleft palate patients require a multidisciplinary approach for optimal care. Cleft palates often occur without known genetic etiology, but they can be part of an inherited syndrome. Otitis media, feeding difficulties, and hearing loss are frequent problems for the cleft palate infant, regardless of the etiology.

Airway problems may occur, particularly with Pierre Robin sequence (which occurs sporadically, as well as in the hereditary Stickler syndrome).

Cleft Lip and Cleft Palate Syndromes

Cleft Palate

Cleft palate (CP) as an isolated malformation is a clinical entity distinct from cleft lip (with or without cleft palate). In isolated CP the risk of recurrence in subsequently born children is about

- 2% if one child has CP
- 6% if one parent has CP
- 15% if one parent and one child have CP

Cleft palate patients have poor eustachian tube function; there is a near universal rate of otitis media before cleft closure and a high rate of middle ear disease throughout life.

22q11 Deletion Syndromes

Gene map locus: 22q11

Autosomal dominant

Fluorescence in situ hybridization (FISH) is a cytogenetic test that involves the use of fluorescent DNA probes to identify small chromosomal abnormalities missed by a previous method. Since the introduction of FISH, it has been possible to show that six clinical syndromes with overlapping manifestations, including learning disabilities, palatal abnormalities, characteristic facial features, and heart defects, all represent manifestations of the 22q11 deletion syndrome.

DiGeorge's Syndrome

Deletions of 22q11 are characteristic in DiGeorge's syndrome. The syndrome is ordinarily sporadic and occurs from a new mutation. DiGeorge patients have

- Parathyroid hypoplasia with absent C cells and hypocalcemia
- Thymic hypoplasia with T-cell immunodeficiencies (decreased CD4 cells)
- Outflow tract defects of the heart, including Fallot's teratology
- Cleft palate (overt or submucous cleft is frequent)

Sensorineural hearing loss with profound deafness can occur. Temporal bone anomalies include bony fusion of the ossicles and the otic capsule.

Velocardiofacial Syndrome (Shprintzen's Syndrome)

Although 22q11 deletions occur in this syndrome as well, velocardiofacial syndrome (VCS) is much less severe than DiGeorge's syndrome. Features of VCS include

- Characteristic facial appearance, including a square nasal tip and longish face
- Velopalatal insufficiency, usually in association with submucous or overt cleft palate (early pharyngeal flap is indicated if an initial trial of speech therapy fails.)
- Otitis media (nearly universal)
- Cardiac defects, often including ventricular septal defect

A main clinical feature of this syndrome is learning disability or mental retardation (usually mild). Measures to improve hearing loss in early childhood (such as myringotomy and tubes) may help optimize achievement.

Stickler Syndrome

Gene map loci: 12q13.11–q13.2, 1p21, 6p21.3

Genes: *COL2A1, COL11A1, COL11A2*

Autosomal dominant

The classic phenotype of Stickler's syndrome is associated with mutations in *COL2A1*, a fibrillar collagen that is arrayed in quarter-staggered fashion to form fibers similar to those of collagen I. Stickler's syndrome includes cleft palate, progressive myopia, a high rate of retinal detachment, radiographic bone abnormalities, and early osteoarthritis. Newborns present with Pierre Robin sequence; other manifestations are not obvious early on, but may be revealed with ophthalmologic and radiological consultation. Features of Pierre Robin sequence include glossoptosis, secondary (soft palate) cleft, and micrognathia. This sequence occurs during fetal development; the tongue falls back, preventing closure of the palatal shelves, and fails to push the mandible forward. Tracheotomy or lip–tongue adhesion may be required for improvement of the airway.

Craniofacial Synostosis

Includes Crouzon syndrome, Apert syndrome, and Pfeiffer syndrome

Gene map locus: 10q26

Gene: *FGR2*

Autosomal dominant

Crouzon syndrome, Apert syndrome, and Pfeiffer syndrome are the three most common types of craniofacial synostosis; each involves premature fusion of at least one of the cranial sutures, with resulting skull shape deformity and midface retrusion with characteristic facies.

Facial appearance is similar in these three syndromes. Appearance of the hands differs. In Crouzon syndrome, the hands are normal. In Pfeiffer's syndrome, there is a broad, square-tipped first digit (finger and toe). In Apert's syndrome, there is syndactyly of the hands and feet. Obstructive sleep apnea syndrome or impaired growth because of upper airway obstruction is common. Tracheotomy may be required. Hearing impairment because of conductive or mixed losses is also common.

These three conditions are all caused by mutations in the same fibroblast growth receptor gene, *FGR2*. Fibroblast growth factor receptors are a family of tyrosine kinase receptors. Mutations in *FGR2* are usually new, rather than inherited, and cases of craniofacial synostosis appear most often with no previous family history. Once the mutation occurs, it is transmitted in an autosomal dominant pattern of inheritance.

Crouzon Syndrome (Craniofacial Dysostosis, Type I)

Features of Crouzon syndrome include hypertelorism, parrot beaked nasal deformity, external strabismus, and midface retrusion. Upper airway obstruction and conductive hearing loss are common. Progressive hydrocephalus is frequent, but intelligence is normal.

Pfeiffer Syndrome

In Pfeiffer syndrome, there is a broad, square-tipped first digit (finger and toe) and facial features similar to Crouzon syndrome. Subtypes of Pfeiffer's syndrome with additional limb anomalies and airway malformations (including tracheal sleeves with stenosis) occur uncommonly. There is a particularly severe variant with cloverleaf skull. This subtype has a high mortality.

Apert Syndrome

In Apert's syndrome, there is syndactyly of the hands and feet. Varying degrees of mental deficiency occur (not characteristic of Crouzon syndrome or Pfeiffer's syndrome), but many individuals do have normal intelligence.

Treacher Collins Syndrome, Mandibulofacial Dysostosis

Gene map locus: 5q32–q33.1 and others (several loci)

Gene: *TREACLE*

Autosomal dominant

There is variable penetrance in Treacher Collins syndrome (TCS), with some patients appearing mildly unusual and others severely affected with facial and otologic deformities. Familial TCS has been mapped to the chromosome 5q32–q33.1 region. The *TREACLE* gene has been cloned from this locus; its function is unknown. Features of TCS variably include antimongoloid slant of the eyes, coloboma of the lower lid, micrognathia, microtia, hypoplastic zygomatic arches, and external auditory atresia (frequently bilateral). Hearing loss and airway obstruction are the primary functional problems.

Infants with bilateral atresia require immediate bone conduction amplification. Reconstructive atresia surgery is sometimes possible, but unfavorable anatomical features usually preclude a good result. The bone anchored hearing aid is the alternative to external/middle ear surgery (see Chapter 31).

Osteopetrosis (Albers-Schönberg Disease)

Gene map locus: 11q12-q13

Autosomal recessive

This progressive disease of endochondral bone causes macrocephaly, deafness, blindness, hepatosplenomegaly, and severe anemia. Symptoms begin in infancy. Bone marrow transplantation or high-dose calcitriol therapy may ameliorate the progression of disease. Deafness and blindness (as well as other cranial nerve deficits) result from bony entrapment of nerves.

HEREDITARY HEARING IMPAIRMENT

Hereditary deafness has long been known to account for at least half the cases of childhood sensorineural hearing loss (SNHL). Among school children, one child in 650 to 2000 has some form of hereditary deafness. Nearly 30% of all cases are associated with additional phenotypic features (syndromic deafness), the presence of which makes the unequivocal diagnosis of hereditary hearing impairment easier. However, the most common cause of deafness in children is nonsyndromic hearing loss; that is, hearing loss presents alone without any other recognizable phenotypic manifestation.

Application of a standard protocol to the evaluation of all hearing-impaired persons is not recommended. Historical information and clinical tests should be tailored to specific age groups and types of hearing loss (**Table 19–1**). Historical information should include a detailed gestational, perinatal, and family history. The family history should explore not only hearing loss but also other family traits, such as night blindness,

Table 19–1 Tests for evaluating the cause of sensorineural hearing loss by age of onset

Type of Evaluation	Neonate/Infant	Child/Adolescent	Adult
Acquired versus hereditary	Serum IgM titers for CMV, rubella, toxoplasmosis, Viral cultures (urine, throat) VDRL/FTA-ABS	Serum IgM titers for mumps VDRL/FTA-ABS Autoimmune panel	VDRL/FTA-ABS Autoimmune panel Serum IgM titers for mumps
Karyotyping	Dysmorphic features, not a known syndrome	Dysmorphic features, not a known syndrome Mental retardation	
CT of temporal bone	Craniofacial syndromes Pendred's and Stickler's syndromes Dilated vestibular aqueduct (DVA)	Progressive SNHL DVA, Pendred's, Alport's, and Stickler's syndromes Cochlear otosclerosis; Stapes gusher syndrome	Cochlear otosclerosis
MRI of IAC		NF2	Vestibular schwannoma, NF2
EKG	Family history of sudden death, syncope, long QT (Jervell and Lange-Nielsen syndrome)	Family history of sudden death, syncope, long QT	
Ophthalmologic tests	Funduscopy for CMV retinopathy	Funduscopy or ERG for retinitis pigmentosa (Usher's syndrome); myopia and retinal detachment (Stickler's syndrome and Norrie's disease); cataracts (NF2)	Funduscopy or ERG for retinitis pigmentosa (Usher's syndrome)
Thyroid tests	Thyroid function tests Perchlorate discharge test (Pendred's syndrome)	Thyroid function tests Perchlorate discharge test (Pendred's syndrome)	
Renal evaluation	Ultrasound (BOR)	Ultrasound (BOR) Urine blood/protein (Alport's syndrome) Urine dermatan sulfate, heparine sulfate (mucopolysaccharidoses)	Urine blood/protein (Alport's syndrome)
Neurological evaluation		Mental retardation	Ataxia, neuropathies (FSH muscular dystrophy, Refsum's disease)
Dental evaluation		Dentinogenesis imperfecta (osteogenesis imperfecta)	
Vestibular testing	Delayed motor development, rotary chair (Pendred's and Usher's syndromes, DFNB4, DVA)	ENG, rotary chair, posturography (i.e., Pendred, Usher, DFNB4, DVA)	ENG, rotary chair, posturography (i.e., NF2, Pendred's and Usher's syndromes, DFNA9)
Molecular testing	NSHL: DFNB1 and mtDNA mutations A1555G, A7445G SHL: disease specific (BOR, Usher's syndrome, etc.) DVA: Pendred's syndrome, DFNB4	NSHL: DFNB1 and mtDNA mutations A1555G, A7445G SHL: disease specific (BOR, Usher's syndrome, etc.) DVA: Pendred's syndrome, DFNB4	Mitochondrial syndromes DFNA6/14 (low-frequency SNHL)

BOR, brancio-oto-renal syndrome; CMV, cytomegalovirus; CT, computed tomography; DFN, TK; EKG, electrocardiogram; ENG, electronystagmography; ERG, electroretinogram; FSH, follicle stimulating hormone; IAC, internal auditory canal; IgM, immunoglobulin M; MRI, magnetic resonance imaging; mtDNA, mitochondrial DNA; NF2, neurofibromatosis type 2; NSHL: nonsyndromic hearing loss; QT, TK; SHL, syndromic hearing loss; SNHL, sensorineural hearing loss; VDRL/FTA-ABS, TK premature graying of hair, fainting spells, kidney abnormalities, goiter, dysmorphic features, and parental consanguinity. Careful evaluation of both hearing-impaired and normal-hearing family members facilitates the diagnosis of most syndromic forms. The examination is supplemented with specific laboratory tests and radiographs based on the clinical assessment. Chromosomal karyotyping is indicated in a child born with an autosomal dominant syndrome whose parents and grandparents show no obvious phenotypic features of the syndrome, especially if there is a history of numerous miscarriages and the phenotype is not recognizable as a previously reported syndrome.

The individual with nonsyndromic deafness presents with hearing loss alone. The different genes mapped in families with nonsyndromic deafness have been named DFN (for deafness) and numbered sequentially as each was discovered. The letter *A* stands for autosomal dominant inheritance, the letter *B* for autosomal recessive inheritance. When nonsyndromic hearing loss is classified by mode of inheritance, roughly 28% is autosomal dominant (DFNA), 78% autosomal recessive (DFNB), and 2% other (X-linked or DFN, mitochondrial inheritance). In general terms, autosomal recessive deafness has a prelingual onset, and autosomal dominant hearing loss is postlingual and progressive. Nonsyndromic hereditary hearing loss is genetically heterogeneous; over 100 genes are predicted to cause hearing loss in humans. To date, over 80 gene loci have been reported: 51 are for DFNA forms, 31 for DFNB forms, six for DFN forms, and two linked to the mitochondrial genome. Just over 30 genes have been identified so far (see Petit et al, 2001, in Suggested Readings for additional information).

The basic classification of deafness has become subtly altered by new genetic information. Traditionally, syndromic and nonsyndromic forms have been separated on the basis of associated clinical features. It is clear that in some cases the identical gene can cause both a syndromic and a nonsyndromic form of hearing loss (e.g., Pendred syndrome and DFNB4). Another major branch point in the classification of deafness has been by etiology, either genetic or acquired causes. In some cases a genetic abnormality and an environmental factor may have combined to cause deafness (e.g., mitochondrial DNA mutations and increased risk of aminoglycoside ototoxicity).

Branchio-oto-renal Syndrome

Gene map locus: 8q13.3

Gene: *EYA1*

Autosomal dominant

This variable syndrome is one of the most common hereditary causes of hearing loss. Although transmission is autosomal dominant, appearance is so variable that the diagnosis is easily overlooked in parents. The gene is a human homologue of the *Drosophila* "eyes absent" gene (EYA) and is named *EYA1*. In the fruit fly, mutations of this gene cause eyeless flies. In the mouse, *EYA1* plays a role in the development of all components of the inner ear, from the emergence of the otic placode, and in the development of cells surrounding the ureteric branches in the kidney. Mutations in the *SIX1* gene also play a role in the molecular basis of branchio-oto-renal syndrome by disrupting *EYA1-SIX1* DNA complexes.

Hearing loss is often mixed, and the sensorineural component is often progressive. Temporal bone pathology and imaging studies show ossicular malformations, dilated vestibular aqueduct, and Mondini-type deformity.

Syndromic features include anomalies of the auricles and branchial clefts (preauricular pits, branchial cleft cysts and fistulas, and unusually shaped auricles). In some families, there are only hearing and head and neck manifestations; in other families, malformations of the renal collecting system and polycystic kidneys sometimes also occur.

Pendred Syndrome

Gene map locus: 7q31

Gene: *SLC26A4*

Autosomal recessive

Pendred syndrome is one of the most common forms of syndromic deafness, and it is characterized by SNHL, cochlea displasia, and euthyroid goiter. The *PDS* gene produces a transporter protein for iodide and chloride. In the mouse inner ear, this protein is expressed in the epithelial cells of the endolymphatic sac and duct, and it is thought to play an important role in inner ear fluid homeostasis.

Hearing loss is usually congenital, sensorineural, bilateral, and severe. In less severe cases, there may be slight progression during early childhood. A widened vestibular aqueduct and an enlarged endolymphatic sac were found in all of a series of 20 Pendred patients in England undergoing appropriate imaging studies. A few of these patients also showed a frank Mondini deformity. Similarly, the vestibular function is often depressed.

Thyroid enlargement can occasionally be present at birth, but it usually appears before the age of 5 years. The thyroid abnormality consists of a mild type of iodine organification defect, demonstrated by a partial discharge of iodide (20–50%) when intravenous perchlorate is given. The perchlorate discharge test alone is nonspecific, but particularly suggestive in a patient with temporal bone computed tomographic (CT) findings.

Mutations in this gene also have been found in a form of nonsyndromic deafness and in enlarged vestibular aqueduct syndrome (see DFNB4).

Alport Syndrome

Gene map loci: Xq22, 2q36–q37

Gene (Xq22): *COL4A5*

Gene (2q36-q37): *COL4A3, COL4A4*

X-linked or autosomal recessive

This syndrome is characterized by progressive nephritis with uremia that is often heralded by hematuria in the first years of life ("red diaper"). Renal insufficiency ultimately develops. The renal lesion is a glomerulitis; red cell casts often accompany the gross or microscopic hematuria.

Approximately 50% of patients have progressive SNHL, although the degree of impairment varies widely. SNHL starts later than the renal manifestations. There have been reports of improved hearing thresholds with therapy to improve renal status. Histopathological findings include degeneration of the stria vascularis, spiral ganglia, hair cells, and endolymphatic hydrops. Collagen types $4\alpha3$, $4\alpha4$, and $4\alpha5$ are found in the basilar membrane, parts of the spiral ligament, and stria vascularis. The X-linked phenotype has been shown to be the result of mutation in the gene for the α-S chain of basement membrane collagen (*COL4A5*). Although the mechanism of hearing loss is not known, loss of integrity of the basement membrane in the spiral sulcus might affect adhesion of the tectorial membrane, and in the basilar membrane and its junction with the spiral ligament, translation of mechanical energy may be affected.

Jervell and Lange-Nielsen Syndrome

Gene map locus: 11p15.5, 21q22.l–q22.2, and others (heterogeneous)

Gene (11p15.5): *KVLQT1*

Gene (21q22.1–q22.2): *KCNE1*

Autosomal recessive

The molecular abnormality in Jervell and Lange-Nielsen syndrome (JLNS) is in potassium ion channels. Endolymph homeostasis is controlled in part by the delayed rectifier potassium channel. Mutations can occur in one or two genes coding for potassium channel proteins. JLNS patients have congenital, severe-to-profound SNHL, prolongation of the QT interval, torsade de pointe arrhythmias (turning of the points, in reference to the apparent alternating positive and negative QRS complexes), and sudden syncopal episodes.

The electrocardiogram in childhood can be initially normal. There is often a family (or personal) history of syncope, or sudden death in a sibling. The cardiac arrhythmia known as torsade de pointe is responsible for mortality. Prompt antiarrhythmic treatment has reduced the high mortality associated with this disease. Patients may be candidates for an automatic implantable defibrillator.

Norrie Disease

Gene map locus: Xp11.4

Gene: *Norrin*

X-linked

Features of Norrie's disease include

- Progressive SNHL in less than half of affected persons
- Specific ophthalmalogic findings (pseudotumor of the retina, retinal hyperplasia, cataracts, etc.). Ophthalmology consult can help make this diagnosis in a male with abnormal mental status or mental retardation presenting with progressive SNHL.
- Mental impairment

Usher Syndrome

This disorder is the most common autosomal recessive syndromic form of childhood SNHL. It affects approximately one half of the 16,000 deaf-blind persons in the United States and 3 to 10% of the congenitally deaf population. Usher's syndrome is characterized by hearing impairment and retinitis pigmentosa (RP). The vision loss from RP is progressive, heralded by complaints of night blindness or visual field defects. Ophthalmologic consultation is helpful, but it may not reveal RP until years after onset of SNHL, unless an electroretinogram is performed. In addition to aural habilitation, because blindness ultimately develops, early intervention should include visual habilitation (e.g., Braille). DNA tests are available and can be used to avoid invasive tests (electroretinogram) in children with suspected Usher's syndrome (congenital deafness and delayed gross motor development from vestibular dysfunction).

Usher syndrome is classified into three different types on the basis of clinical findings.

Usher Type 1

Gene map loci: l1q13.5, 11p15.1, but highly heterogeneous, with seven different loci

Gene: *MYO7A, USH1C,* and others

Type 1 is the most common and most severe type of Usher syndrome. It is genetically heterogeneous and has seven subtypes (1A–1G). Clinically, type 1 features

- Profound congenital deafness with absent vestibular function. Children typically do not walk until more than 1 year of age and have balance difficulties: clumsiness, frequent falling, and difficulty swimming (because of poor vestibular function).
- Progressive blindness with onset of RP by age 10 years

Mutations in the *MYO7A* gene occur in type 1B and account for 75% of cases of Usher syndrome type 1. Mutations in the *MYO7A* gene also account for other nonsyndromic forms of SNHL (DFNB2 and DFNA11). Myosin 7A is an unconventional motor molecule present in the rods and cones of the retina, and in the cochlear and vestibular sensory hair cells during both human and mouse embryonic development. This motor protein is present in the stereocilia and near the hair cell–supportive cell junction; in mice, myosin 7A plays a role in forming and stabilizing the hair bundles. The Shaker-1 mouse is an animal model of Usher's syndrome type 1 with a similar genetic mutation, similar histological appearance of the degeneration of the organ of Corti, and deficits in both hearing and vestibular function. This mouse model facilitated mapping of the gene in humans.

Usher Type 2

Gene map locus: 1q41 and others

Gene: *USH2A*

Type 2 is less common and less severe than type 1. Clinical features include

- Moderate to severe congenital deafness, with typically sloping audiogram, and normal vestibular function. Types 1 and 2 can be differentiated by tests of vestibular function.
- Onset of RP in late teens

Usher Type 3

Gene map locus: 3q21–3q25

Gene: *USH3*

Type 3 is the least common form of Usher's syndrome and has only recently been confirmed as a distinct entity. The SNHL, vestibular dysfunction, and RP are variable and progressive.

Waardenburg Syndrome

Autosomal dominant

Waardenburg's is the most commonly recognized syndromic deafness. There is variable expression: hearing impairment is present in no more than half of the affected patients. Syndromic features other than hearing loss should be sought in the family history of newly diagnosed cases of SNHL to rule out the syndrome. There are two major types of Waardenburgs that together account for almost all cases. Abnormal migration of neural crest cells occurs.

Type 1

Gene map locus: 2q35

Autosomal dominant

The one invariable feature of Waardenburg type 1 is dystopia canthorum: apparent broad spacing of the eyes due to lateral displacement of the medial canthi. This appearance is sometimes described as a "broad nasal root." True hypertelorism is not present: the interpupillary distance is normal.

Spotty pigmentation frequently occurs along with dystopia canthorum but is not necessary for diagnosis of type 1 and is not always present. White forelock is striking but unusual; much more common are patches of depigmented skin, patches of gray or white hair (including eyelashes), or very premature grayness. The eyes may each be of a different hue (heterochromia iridis), or there may be light sapphire blue eyes in a darkly pigmented person.

SNHL is present in only 20% of Waardenburg syndrome type 1 patients, and it may be unilateral or mild enough to be unapparent on family history. SNHL is often progressive. *PAX3* mutations are responsible for most type 1 cases.

The Splotch mouse is an animal model of Waardenburg type 1. The use of this mouse model, in conjunction with human studies, facilitated gene mapping.

Type 2

Type 2 is characterized by pigmentary abnormalities without dystopia canthorum. Although no dystopia canthorum is present, patchy pigmentation occurs (e.g., heterochromia iridis, white skin patches and/or white hair patches, or premature grayness). Hearing loss is more common than in type 1; 50% of Waardenburg's type 2 patients have SNHL. Again, SNHL is frequently progressive and may be unilateral.

Mutations in *MITF* and *PAX3* (both genes code for transcription factors) are responsible for Waardenburg syndrome type 2.

Genetic Causes of Conductive (and Mixed) Hearing Loss

Otosclerosis Otospongiosis

Gene map locus: 15q26.1–q territory

Otosclerosis features include isolated endochondral bone sclerosis of the labyrinthine capsule, commonly resulting in the fixation of the stapediovestibular joint of the oval window. Stapes surgery can improve hearing thresholds by as much as 55 dB. Other parts of the temporal bone can be affected. Nearly 10% of affected persons develop inner ear deafness (cochlear otosclerosis) for which there is no cure, but drugs such as disphosphonates and calcium-fluoride compounds have been used to slow down the progression of sensorineural hearing loss.

- Clinical otosclerosis has a prevalence of 0.2 to 1% among Caucasian adults and occurs approximately twice as frequently in females as in males. Segregation analysis performed in relatives and patients with familial otosclerosis indicates that a rare dominant gene is involved along with an extensive polygenic component. Less than 20% of affected patients in these families carry the dominant gene. Sporadic cases would be expected to have an even lower incidence of this gene. Penetrance varies according to age and sex. One otosclerosis gene has been mapped to chromosome 15.

Mixed Deafness with Stapes Fixation and Perilymphatic Gusher (DFN3)

Gene map locus: Xq21.1

X-linked

The mutant gene in DFN3 is in a transcription factor with a POU domain known as brain-4.

This X-linked mixed type of deafness is characterized by the following:

- Expression is mostly in males (homozygotes) who exhibit mixed hearing loss with a progressive SNHL component and impaired vestibular function.
- The conductive component is secondary to a congenitally fixed stapes. Profound sensorineural deafness can mask the conductive component.
- Perilymphatic gusher, with postoperative drop in sensorineural thresholds, occurs with stapes surgery, which is contraindicated.
- Dilatation of the lateral end of the internal auditory canal is a frequent finding on temporal bone imaging. Preoperative CT imaging in young men with progressive mixed loss (particularly with maternal history of hearing loss) is indicated to avoid stapes surgery in DFN3 patients (**Fig. 19–3**).
- Mothers of affected sons (obligate heterozygotes) have mild SNHL and normal vestibular function.

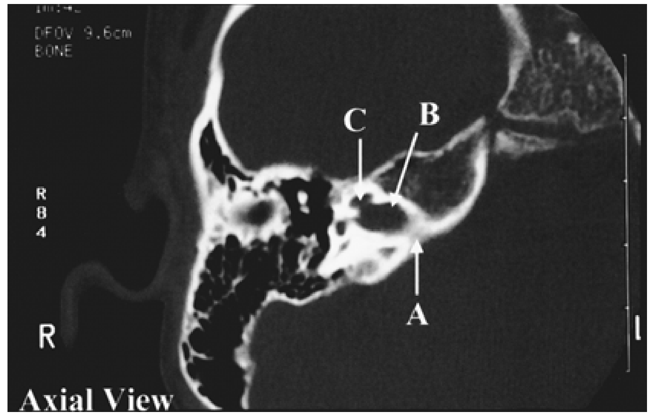


Figure 19–3 *A high-resolution computed tomographic scan of a patient's inner ear at the modiolar level showing dilation of the internal auditory canal (IAC, labeled B), hypoplasia of the basal turn of the cochlea with no detectable bony modiolus (C) providing a route of direct communication between the subarachnoid space of the posterior cranial fossa (A) and the perilymphatic space of the inner ear. Stapes fenestration surgery is contraindicated in this patient due to the high risk of hearing loss from a cerebrospinal fluid "gusher."*

Osteogenesis Imperfecta Type 1 (Osteogenesis Imperfecta with Blue Sclerae)

Chromosomal locus: 7q22.1 and others

Autosomal dominant

There are multiple forms of osteogenesis imperfecta (OI), some of which are clinically mild. Many involve mixed hearing loss and stapes fixation. OI1 features:

- Bone fragility: fractures are rare in the neonatal period, but fracture tendency (even with minimal trauma) is constant from childhood to puberty, and decreases thereafter until late life.
- Blue sclerae that remain blue throughout life
- Abnormal collagen type 1 (in most cases)

Conductive or mixed hearing loss occurs in ~50% of families; beginning in the late teens the SNHL gradually progresses to profound deafness, associated with tinnitus and vertigo by age 50 to 60 years. Conductive hearing loss is secondary to a fixed stapes, often with fractured crura and a thickened footplate.

Stapes surgery is helpful for closure of the air–bone gap. Stapes surgery in OI patients yields the same functional results as would be expected in otosclerosis patients, even though the underlying etiology is different.

Otitis Media

Otitis media is a spectrum of diseases with a high prevalence in the population; thus an inherited susceptibility for otitis media is not readily apparent. Genetic factors can modulate the host's response to infectious agents and the development of disease. Epidemiological data and twin studies have shown that heritable factors affect the individual susceptibility to both recurrent acute and chronic serous otitis media. Furthermore, it has been reported that the incidence of otitis media is significantly lower in patients with otosclerosis, an autosomal dominant hereditary condition with low penetrance, suggesting that the genetic factors associated with otosclerosis infer a protective effect. The genetic mechanism of resistance and/or susceptibility to otitis media is not yet known.

Nonsyndromic Sensorineural Hearing Loss

SNHL in a child or young adult without known acquired cause and without any associated unusual physical features is called nonsyndromic deafness. Most nonsyndromic forms of deafness are monogenic and rare, with the exception of one form caused by inheritance of a homozygous autosomal recessive mutant gene for the gap junction protein connexin 26. Genetic heterogeneity and incomplete knowledge of most forms of nonsyndromic SNHL are important challenges to understanding the molecular pathogenesis of nonsyndromic SNHL; however, a classification scheme based on the primary gene defect has been proposed. **Fig. 19–4** shows some of the genes involved in nonsyndromic hereditary hearing loss according to their effect on inner ear structures: hair cell, supportive cell, and tectorial membrane.

Connexin 26 (CX26): DFNB1 and DFNA3

Gene map locus: l3ql l–q12

Gene: *GJB2*

Mutations in this gap junction protein gene account for the highest known percentage of nonsyndromic congenital SNHL, estimated between 10 and 80% of nonsyndromic SNHL cases, depending on the reference population.

Mutations in this gene can cause either autosomal dominant (DFNA3) or autosomal recessive (DFNB1) deafness, depending on the exact change in the nucleotides. Most cases are recessive.

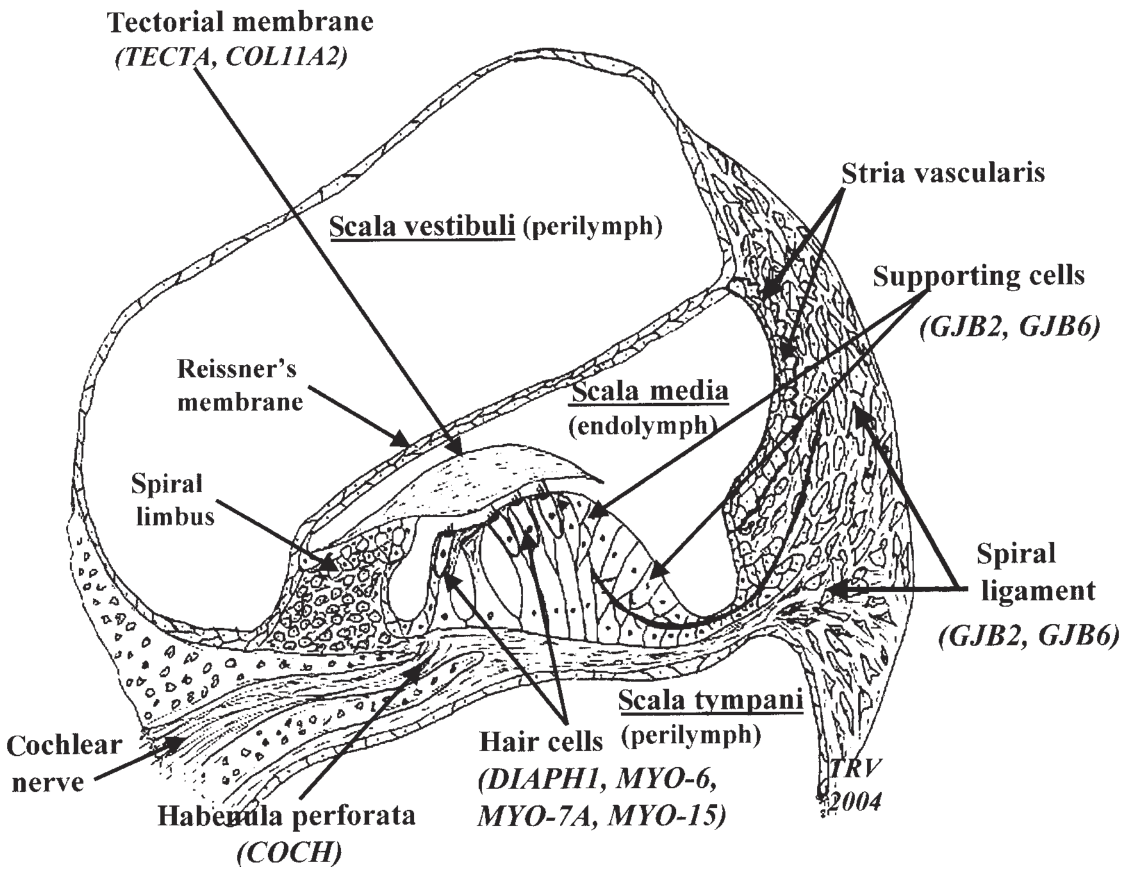
DFNB1 is characterized by congenital, nonprogressive, mild to profound sensorineural hearing loss, without other associated features, and with normal vestibular function. The severity of the hearing loss is variable and may differ between siblings and other close relatives sharing the same mutation.

CT scans of the temporal bone show a normal otic capsule.

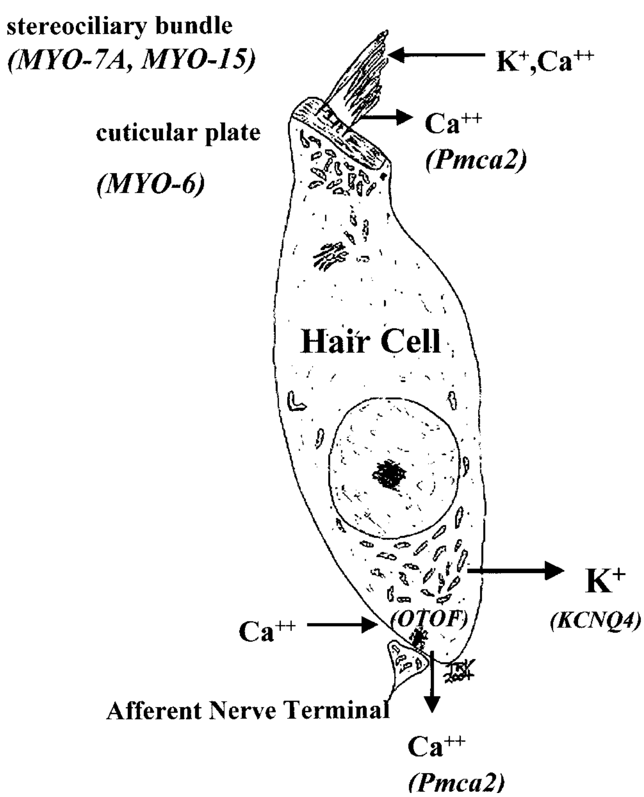
Six connexin 26 molecules oligomerize into hexamers called connexons, and two connexons, one from each cell, join in the extracellular gap to form a transmembrane gap junction channel that allows intercellular flow of ions and molecules between cells. In the inner ear, gap junctions allow synchronization of electrical activity in excitable tissues and the maintenance of the endolymphatic potential.

DFNB1 accounts for ~50% of congenital autosomal recessive deafness in the United States, Europe, and Australia. Its approximate prevalence in the general population is 14 per 100,000. Numerous deafness-causing mutations have been identified, but by far the most common one is *35delG*. The *35delG* mutation is estimated to be responsible for 10% of all childhood deafness, and 20% of all childhood hereditary hearing loss. Other recessive mutations are common in some ethnic groups. The *167delT* mutation, rare in most groups, is prevalent in Ashkenazi Jewish families with nonsyndromic recessive deafness. In the Ashkenazi Jewish population the carrier rate for these two connexin 26 deletion mutations (*35delG* and *l67delT*) is nearly 5%. This carrier frequency predicts a prevalence of one deaf individual in ~1800 and appears to account for most cases of nonsyndromic recessive deafness in patients with a Jewish background.

Molecular genetic testing of DFNB1 has become widely available and can identify 95% of deafness-causing mutations. However, ~10% of patients with connexin 26 gene mutations have only one mutated allele, and some families in whom the diagnosis of DFNB1 was established by linkage studies have no connexin 26 mutation; thus failure to detect a connexin 26 mutation does not exclude the diagnosis of DFNB1. The DFNB1 chromosomal locus contains two functionally related connexin genes, *GJB2* (connexin 26) and *GJB6* (connexin 30), and DFNB1 may result from a monogenic or digenic pattern of inheritance.



A



B

Figure 19–4 *Schematic presentations of the probable sites of action of gene mutations that are representative of genes that underlie isolated nonsyndromic deafness within the* ***(A)*** *cochlea and* ***(B)*** *auditory hair cells. TECTA, COL11A2: genes encoding for a major noncollagenous and a collagenous extracellular matrix component of the tectorial membrane, involved in DFNA8/12 and DFNA13, respectively. GJB2, GJB6: genes encoding for gap junction proteins, involved in DFNB1 and DFNA3. COCH: a gene whose product is detected in cells that accompany the distal axons of the auditory neurons at the level of the habenula perforata, involved in DFNA9. MYO-7A: a gene that encodes for a motor protein expressed in auditory hair cells, involved in DFNB2, DFNA11, and Usher type 1. MYO-6: a gene that encodes for a motor protein expressed in the cuticular plates of auditory hair cells, involved in DFNA22. MYO-15: a gene that encodes for a motor protein expressed in the stereociliary bundles and cuticular plates of auditory hair cells, involved in DFNB3. KCNQ4: a gene that encodes for a K^+ channel subunit, involved in DFNA2. OTOF: a gene that encodes for a vesicle trafficking protein, involved in DFNB9. Pmca2: a mouse gene that encodes for a Ca^{2+} pump component (not yet implicated in human deafness).*

DFNA9

Gene map locus: 14q12–q13

Gene: *COCH*

Autosomal dominant

DFNA9 patients show onset of hearing loss in young adulthood (20–30 years of age). Impairment is initially worse at high frequencies, and then variably progresses to anacusis by 40 to 50 years of age. Clinical vestibular symptoms mimicking Meniere's disease occur.

The *COCH* gene protein cochlin was detected in cells accompanying neurons at the habenula perforata. Acidophilic deposits, consistent with mucopolysaccharide ground substance, have been noted at that anatomical site in temporal bones from DFNA9 patients

DFNB3

Gene map locus: l7p11.2

Gene: *MYO15A*

Autosomal recessive

- DFNB3 is a profound, congenital, nonsyndromic SNHL.
- Animal model: mouse mutant shaker-2', showing mutations of *MYO15*, another unconventional myosin gene

DFNB4 (See Also Pendred Syndrome)

Gene map locus: 7q31

Gene: *PDS*

Autosomal recessive

The *DFN4* gene is apparently the same gene that causes Pendred's syndrome, but in this nonsyndromic recessive deafness there are no associated thyroid abnormalities on physical exam or thyroid function tests. SNHL is usually progressive, and CT scan of the temporal bone shows dilated vestibular aqueduct/Mondini deformity.

Nonsyndromic Deafness and Abnormal Myosin 7A Gene (See Also Usher Syndrome)

Gene map locus: 11q13.5

Gene: *MYO7A*

Both recessive (DFNB2) and dominant (DFNA11) forms of nonsyndromic deafness have been found in different parts of the world in families with myosin 7A gene mutations. These nonsyndromic families do not have impairment in vision (recall that myosin 7A is abnormal in Usher's syndrome type 1, the single most common cause of deaf/blindness).

DFNA2

Gene map locus: 1p34

Gene: *KCNQ4*

Autosomal dominant

Families with DFNA2 are characterized by progressive deafness involving preferentially the high frequencies. *KCNQ4* encodes a subunit of a potassium channel of the outer hair cells that play a role in the maintenance of the electrical properties of these cells. *KCNQ4* is the second member of the *KCNQ* gene family to be involved in SNHL, the first being the *KCNQ1* gene responsible for JLNS.

DFNA8/DFNA12

Gene map locus: 11q22–q24

Gene: *TECTA*

Autosomal dominant

Alpha-tectorin is one of the major noncollagen components of the tectorial membrane in the organ of Corti. Mutations in the α-tectorin gene, *TECTA*, are responsible for at least two autosomal nonsyndromic deafness forms.

Mitochondrial DNA Deafness

Hearing loss accompanies many mtDNA disorders, perhaps reflecting the highly metabolic state of the hearing process. Initially, mtDNA defects were described in some systemic neuromuscular disorders, such as Kearns-Sayre syndrome, MERRF, and MELAS, as well as in families with diabetes mellitus and SNHL. More recently, nonsyndromic deafness has been shown to occur with mtDNA mutations, with remarkable heterogeneity of phenotypic expression even with the same genetic defect, suggesting the possibility of a nuclear gene in modifying the effect of the mutation.

The *A1555G* mutation is one of the genes encoding mitochondrial ribosomal RNA (12S rRNA) and has been recognized as the most common cause of aminoglycoside-induced deafness. Although sufficiently high doses of aminoglycosides will cause SNHL in any person, deafness will occur after even small dosages of aminoglycoside antibiotics in patients who carry this mutation. The mitochondrial ribosome in the cochlea is the most likely target of aminoglycoside ototoxicity, because the "natural target" of aminoglycosides is the bacterial ribosome. The mechanism of ototoxicity of aminoglycosides is thought to be reduced production of adenosine triphosphate (ATP) in the mitochondria of hair cells. The *A1555G* mutation also has been described in families with SNHL without aminoglycoside exposure.

Other families with nonsyndromic SNHL and maternal inheritance have been described to carry a mutation at nucleotide 7445 in the mitochondrial transfer RNA *(tRNA)*$^{Ser(UCN)}$ gene. Furthermore, acquired mtDNA defects have been implicated in the aging process and in the development of presbycusis, in combination with other environmental factors.

Genetic Screening and Molecular Diagnosis of Deafness

Genetic screening is defined as the analysis of human DNA to detect heritable-related mutations. A genetic test is one to detect a heritable disease; we will primarily refer to DNA tests for SNHL, but keep in mind that other means to diagnose genetic disease are available, such as RNA, chromosomes, proteins, and certain metabolites.

The recent availability of DNA tests to diagnose genetic hearing loss has revolutionized the way we approach these cases. In some cases, the DNA test has supplanted other, more invasive and less accurate tests. The goal of genetic testing is to establish an etiologic basis for hearing loss in the most efficient manner possible. Based on the results of the clinical evaluation, the following should be considered.

Syndromic forms of hearing loss have a genetic origin, except for congenital rubella, toxoplasmosis, and cytomegalovirus embryopathies. When syndromic hearing loss is suspected, gene-specific testing should be performed. Available DNA tests for diagnosis of syndromic deafness exist for Waardenburg, Usher, Jerwell and Lange Nielsen, and Pendred syndromes. For a complete list of syndromes and corresponding DNA tests, see Geneclinics (http://www.geneclinics.org). The evaluation can be complemented with specific viral titers and cultures (**Table 19–1**).

Nonsyndromic hearing loss (NSHL) is the most common type of genetic deafness, and among these the most common type of NSHL inheritance is autosomal recessive (ARNSHL). A family history is not usually evident in ARNSHL because sporadic cases predominate, but the hearing loss is usually severe and occurs at birth. Autosomal dominant NSHL tends to occur later in life, is progressive, and usually not severe. It is interesting that in some populations, particularly in Europe and the midwestern section of the United States, one gene alone accounts for just over half of all cases of ARNSHL; that is, the connexin 26 gene (*Cx26*). Thus the following are recommendations for genetic screening of NSHL:

- In neonates with congenital hearing loss and no obvious family history: *Cx26* mutation screening by gene sequencing and cytomegalovirus immunoglobulin M (IgM) titers
- The patient has a family history and other first-degree hearing-impaired relative(s): *Cx26* mutation screening and gene-specific mutation screening if the pedigree shows autosomal dominant inheritance
- The pedigree suggests mitochondrial DNA inheritance (maternal inheritance): testing for the *A1555G* mutation (associated with aminoglycoside ototoxicity) and the *A7445G* mutation, after excluding *Cx26* mutations
- If nonsyndromic deafness is suspected and both parents are deaf, *Cx26*-related deafness is strongly suspected; because *Cx26* deafness is the most common in the United States, the vast majority of marriages between deaf individuals who produce deaf offspring are between individuals with *Cx26*-related deafness
- In patients with progressive SNHL, imaging studies are recommended to identify inner ear malformations. If a cochlear dysplasia is found (Mondini deformity, dilated vestibular aqueduct): screening for *SLC26A4* mutations for Pendred syndrome/DFNB4.

After genetic testing it will be possible to ascribe a genetic etiology to the hearing loss in many persons. For example, a child may be diagnosed with *Cx26* deafness if two mutated alleles are found. We then know the cause of the child's deafness with certainty and can accurately predict the chance of recurrence in a subsequent child. Alternatively, the test may be negative. A negative screening test does not mean that the deafness is not genetic. This distinction is subtle but very important and must be conveyed to parents prior to testing. In patients with a negative family history and a negative test for *Cx26*, the probability that the deafness is genetic can be given, and this probability is based on the number of hearing siblings and the ethnic group.

The benefits of genetic testing in single-gene diseases when the genetic loci are known are obvious, such as in DFNB1, and include determination of the cause of hearing loss, avoidance of unnecessary and costly tests, determination of the chance of recurrence of deafness in the family, and identification of relatives at risk. Prenatal diagnosis is possible by obtaining fetal DNA through amniocentesis or chorionic villi sampling, but there are important pitfalls.

Often, there are several loci that cause syndromes or diseases; ruling out one or two may be possible, but it will not comprehensively rule out the possibility of a trait. Prenatal diagnosis is therefore not available for every disease associated with a known mutant gene.

Merely having the mutant gene does not necessarily mean developing the disease. In most known oncogenes, carrying the mutation places an individual at increased risk but does not predict that cancer will actually occur. Therefore, prenatal testing is currently limited for selected conditions where the clinical usefulness of the screening test has been proven; for example, tests with high predictive value in disorders with specific therapeutic interventions to reduce risk in genetically susceptible individuals.

GENETIC COUNSELING

Genetic counseling is the process of providing individuals and families with information on the nature, inheritance, and implications of genetic disorders to help them make informed medical and personal decisions. The purpose of genetic counseling is to ensure that the parents/patients understand the findings and limitations inherent in any genetic test. Parents are usually given detailed counseling before undergoing genetic testing to ensure that they make informed decisions about the use of tests with complex personal implications. Genetic counseling is usually nondirective; that is, counseling provides sufficient information to empower families or individuals to determine the best course of action for them but avoids making recommendations.

Optimally, counseling includes both pre- and post-test sessions. A pretest session focuses on factual information, including an explanation of the disorder, modes of inheritance, and genetic testing options including their risks, benefits, and limitations. A post-test session includes an explanation of test results, an assessment of the psychological impact that the results may have on the parents and child, and a description of treatment and supportive resources that are available for the family needs.

SUMMARY

The otolaryngologist in the era of genomic medicine will have the opportunity to use the patient's own unique genome to determine the optimal management approach (i.e., preventive, diagnostic, and/or therapeutic). Future challenges include the constantly changing knowledge base on concepts such as genetic variability, interaction of an individual's inherited genetic traits with the environment, and clinical usefulness and reliability of available DNA tests, as well as patient privacy and confidentially issues related to genetic testing.

Suggested Readings

Casselbrant ML, Mandel EM, Fall PA, et al. The heritability of otitis media: a twin and triplet study. JAMA 1999;282: 2125–2130

Del Castillo I, Villamar M, Moreno-Pelayo MA, et al. A deletion involving the connexin 30 gene in nonsyndromic hearing impairment. N Engl J Med 2002;346(4):243–249

Fischel-Ghodsian N, Bykhovskaya Y, Taylor K, et al. Temporal bone analysis of patients with presbycusis reveals high frequency of mitochondrial mutations. Hear Res 1997;110: 147–154

Genetic Evaluation of Congenital Hearing Loss Expert Panel. Genetics evaluation guidelines for the etiologic diagnosis of congenital hearing loss. Screening. Genet Med 2002;4(3): 162–171

Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921 [errata, Nature 2001;411:720, 412:565]

Manolidis S, Alford RL, Smith RJH, et al. Do the genes that cause otosclerosis reduce susceptibility to otitis media? Otol Neurotol 2003;24:868–871

McKusick VA. Online Mendelian Inheritance in Man (OMIM). Available at: http://www.ncbi.nlm.nih.gov/Omim/. Accessed March 15, 2004

Petit C, Levilliers J, Hardelin JP. Molecular genetics of hearing loss. Annu Rev Genet 2001;35:589–646

Van Camp G, Smith RJH. Hereditary Hearing Impairment Homepage. Available at: http://www.uia.ac.be/dnalab/hhh/. Accessed March 16, 2004

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. The *p53* gene encodes for a protein that
   - A. Enhances hearing acuity
   - B. Increases tumor metastases
   - C. Arrests the cell cycle at G1 in a cell with damaged DNA
   - D. All of the above

2. Usher's syndrome type 1 affects
   - A. The visual system only
   - B. The auditory system only

C. The vestibular system only
D. All of the above

3. Mutations in which gap junction protein gene account for the highest percentage of nonsyndromic congenital sensorineural hearing loss?
   A. Connexin 43
   B. Connexin 30
   C. Connexin 26
   D. Connexin 32
4. The following are features of hereditary tumors except
   A. Bilateral and multicentric tumors
   B. More aggressive course than nonhereditary tumors
   C. DNA mutations are found in the tumor cells only.
   D. May occur in patients without risk factors
5. The following are examples of the benefits of genetic screening except
   A. A negative screening test for connexin 26 mutations in a child with congenital deafness is proof that the hearing impairment is not genetic.
   B. In relatives of patients with MEN2A and medullary thyroid carcinoma, blood DNA screening for *RET* gene mutations is available to evaluate the need for prophylactic thyroidectomy.
   C. In adult relatives of patients with neurofibromatosis II, magnetic resonance imaging with gadolinium of the brain is indicated even if the patient is asymptomatic.
   D. DNA tests are being used for the diagnosis of some genetic diseases instead of other invasive, costly tests.
6. Autosomal recessive inheritance is characterized by all of the following except
   A. Skipped generations
   B. The chance of disease recurrence increases with each affected child in the family.
   C. 25% chance of disease recurrence in successive children
   D. No sex predilection
   E. Consanguinity

Part II

THE EAR, HEARING, AND BALANCE

Chapter 20

EMBRYOLOGY OF THE OUTER, MIDDLE, AND INNER EAR

Thomas R. Van De Water and Hinrich Staecker

This chapter discusses the germ layer origins and the embryologic mechanisms that act to form the components of the outer, middle, and inner ears and their contribution to the normal development and dysmorphogenesis of these structures.

NORMAL EMBRYOLOGICAL DEVELOPMENT

Outer Ear

The outer ear consists of the pinna, the external auditory canal, and the outer (cuticular) layer of the tympanic membrane. These structures originate from interactions between the surface ectoderm and cephalic mesenchyme (formed by contributions from both the mesoderm and neural crest tissues) of the head area.

The pinna is formed as the neural crest tissues of the first (mandibular) and second (hyoid) branchial arches interact with the overlying surface ectoderm of these arches (**Fig. 20–1**). This occurs in humans between the fourth and sixth weeks of gestation. During this time there are six swellings that occur equally on the first and second arches called the branchial hillocks of His (**Fig. 20–2**). Of importance is that both the cartilaginous framework and connective tissues of the pinna are derived from the neural crest cells of the first and second branchial arches, and the musclature of the pinna is derived from head mesoderm. The musclature of the pinna includes three extrinsic auricular muscles—the posterior auricular muscle (pulls the pinna back), the superior auricular muscle (pulls the pinna up), and the interior auricular muscle (pulls the superior portion of the pinna forward), as well as six intrinsic auricular muscles (responsible for the folds, or plicae, of the pinna). The three extrinisic and six intrinsic muscles of the pinna are innervated by branches of the seventh cranial nerve.

The external auditory canal is formed around the fourth week of gestation by the invagination and medial growth of the first branchial groove that lies between the first and second branchial arches (**Fig. 20–1**). It is at this time that there is a brief interaction between the ectoderm of the ingrowing primary auditory canal and the endoderm of the dorsal division of the first pharyngeal pouch (developing middle ear cavity). This ectodermal–endodermal tissue interaction is transient. The ingrowth of the branchial groove continues, but due to continuing growth and expansion of head mesenchyme, the epithelial walls of the developing external auditory canal are forced together and appear as a solid chord of epithelium (i.e., the meatal plate), which terminates in a disklike

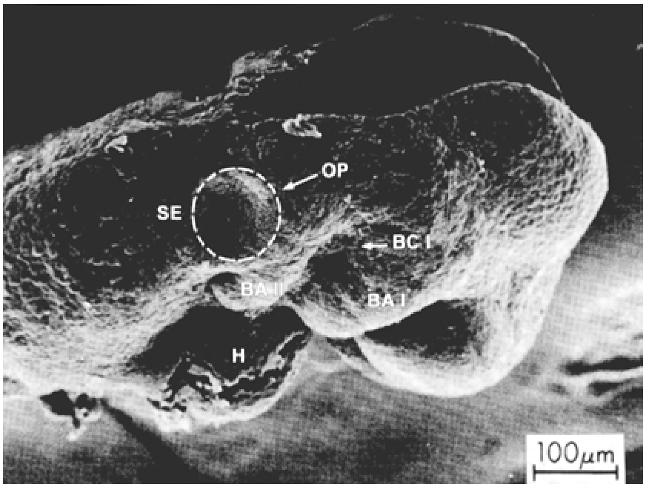


Figure 20–1 *A scanning electron micrograph image of the head region of an 8.5-day-old mouse embryo showing the otic pit (OP), first (BA I) and second (BA II) branchial arches, first branchial cleft (BC I), heart (H), and cephalic surface ectoderm (SE).*

swelling against the ventrolateral wall of the developing middle ear cavity. The formation of this false core of epithelial cells occurs during the eighth week of gestation and represents the secondary (osseous) portion of the external auditory canal. The primary (cartilaginous) portion of this canal forms from the area of the initial ingrowth of the first branchial groove. By the 28th week of gestation the secondary canal opens as the walls of the meatal plate separate, and it becomes the osseous medial extension of the external auditory canal. The disklike swelling that represented the most medial extension of the meatal plate now becomes the external epithelial layer of the tympanic membrane. The tympanic membrane has two other layers, which are discussed in the following middle ear section.

Middle Ear

The middle ear cavity is composed of the middle fibrous and inner mucosal layers of the tympanic membrane, the ossicular chain, the eustachian (pharyngotympanic) tube, and the middle ear cavity.

The other layers of the tympanic membrane are the middle fibrous (pars tensa) layer, which originates from cephalic mesenchymal cells, and the inner epithelial layer, which originates from the endoderm of the middle ear cavity; that is, the epithelial lining of the first pharyngeal pouch. The ring of fibrous cartilage (tympanic annulus) that anchors the tympanic membrane into the sulcus

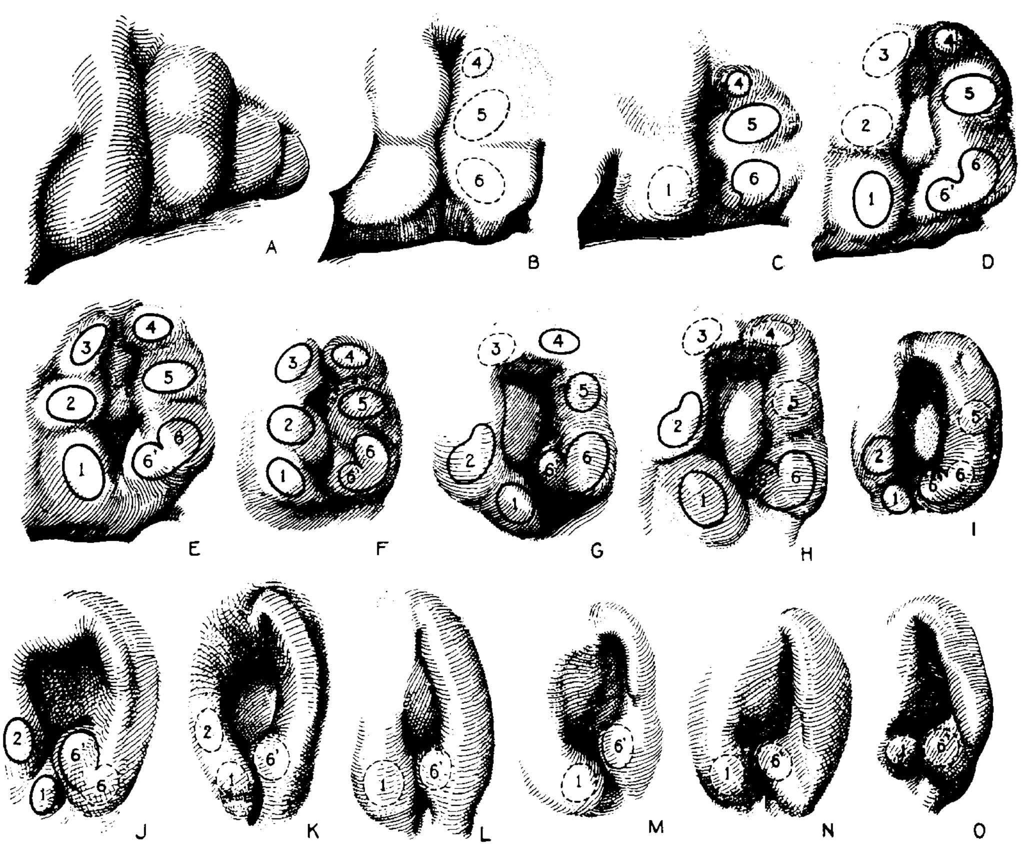


Figure 20–2 ***(A–O)*** *A series of drawings taken from the Carnegie collection of human embryos depicting the formation of pinna from the early branchial hillock stage* ***(A)*** *until the stage of the completion of the primitive pinna* ***(O)****. Numbers 1 to 6 identify the branchial hillocks of His that are hypothesized to contribute to formation of the pinna.*

tympanicus of the middle ear cavity is also derived from the cephalic mesenchyme that forms the middle fibrous layer of the tympanic membrane.

All of the mucosal membranes lining the middle ear cavity and investing the middle ear ossicles originate from the endodermal lining of the first pharyngeal cleft, which begins to expand and form the tubotympanic sulcus at around the fourth week of gestation. This expansion of the first pharyngeal pouch ultimately forms the tympanic cavity, and its proximal connection to the pharynx becomes the eustachian tube. A few mastoid air cells begin to develop during the late fetal period (i.e., 20–38 weeks), but most of the air cells of the mastoid develop postnatally. The process of the endodermal lining of the middle ear cavity invading the mastoid air cells is called pneumatization, and this also is predominantly a postnatal process. The ossicular chain is derived from the most dorsal population of neural crest cells of the first and second branchial arches. The head and neck of the malleus and body, short process, and beginning third of the long process of the incus are derived from the first arch mesenchyme (Meckel's cartilage), and the manubrium of the malleus and the distal two thirds of the long process of the incus originate from the second arch mesenchyme (Reichert's cartilage). The stapes, with the exception of its footplate, is derived from the second arch mesenchyme. The stapedial footplate originates from the same population of cephalic mesoderm cells that forms the otic capsule, which is the precursor of the petrous portion of the temporal bone that encases and protects the membranous labyrinth (**Fig. 20–3**). The ossicles begin to form as mesenchymal condensations in the region of the developing tubotympanic sulcus around the 4th week of gestation and continue to mature through the 38th week of development, when the manubrium of the malleus attaches to the tympanic membrane and the stapedial footplate completes its attachment within the oval window via the annular ligament of the stapes. The ossicles begin the process of ossification around the 10th week of development, and the ossicular chain is functional at birth but not fully mobile until ~2 months after birth. The five ligaments that act to suspend the malleus (i.e., anterior, superior, and external ligaments) and incus (i.e., posterior and superior ligaments) within the middle ear cavity are all derived from the middle ear cephalic mesenchyme. The two skeletal muscles within the middle ear cavity that modulate the vibratory function of the ossicular chain originate from head mesoderm. The tensor tympani muscle, a first branchial arch mesoderm derivative, inserts into the neck region of the malleus and receives innervation from a branch of the otic ganglion of the fifth cranial nerve. The stapedius muscle inserts into the neck of the stapes, receives innervation from the tympanic branch of the seventh cranial nerve, and originates from second branchial arch mesoderm.

Inner Ear

The inner ear consists of the vestibular and auditory divisions of the membranous labyrinth, the auditory (spiral) and vestibular (Scarpa's) ganglia of the eighth cranial nerve, the perilymphatic spaces of the labyrinth, and the bony labyrinth of the temporal bone.

The membranous labyrinth is derived from the otic placode (**Fig. 20–1**), which forms early during the fourth week of gestation out of head ectoderm after a series of sequential inductive interactions between presumptive ear field ectoderm and the inductor tissues of heart mesoderm, chorda mesoderm (notochord), and

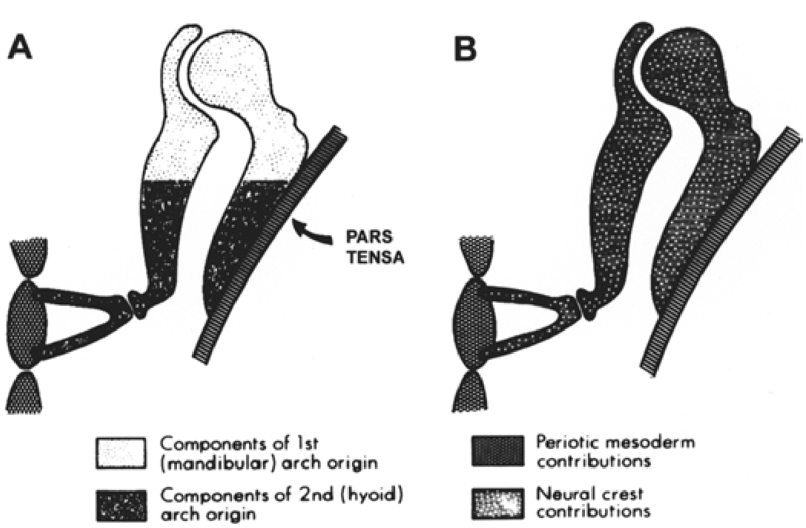


Figure 20–3 *Schematic drawings depicting the theories of the origins of the different components that make up the ossicular chain.* ***(A)*** *The medical genetics–based concept that both BA I and BA II contribute to the formation of the malleus and incus, with BA II forming the superstructure of the stapes.* ***(B)*** *The known contributions of cephalic neural crest cells and periotic mesoderm to the ossicular chain.*

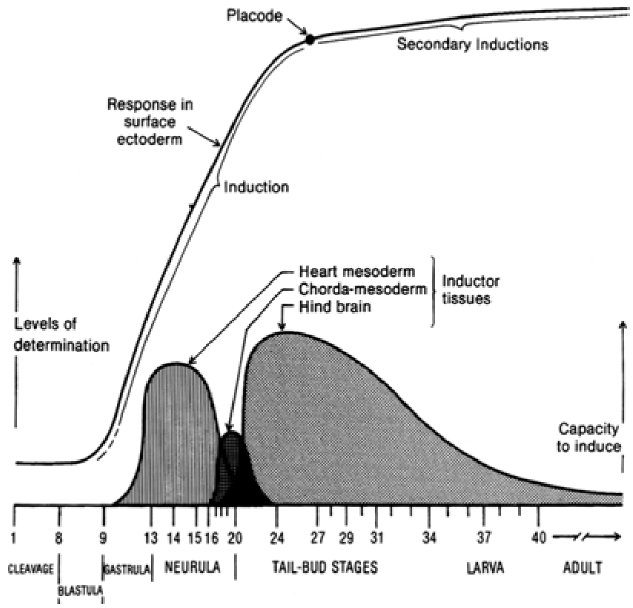


Figure 20–4 *A graphic presentation of the temporal series of tissue interactions between the cephalic surface ectoderm and the underlying primitive mesodermal and neural tissues that induce the formation of the otic placode in an amphibian embryo.*

hind brain (rhombencephalon) (**Figs. 20–4** and **20–5**). By the end of the fourth week of gestation, the otic placode has become a closed vesicle. Neurogenic foci are formed within the otic epithelium and produce the statoacoustic ganglion by the delamination of neuroblasts from the sites of the neurogenic foci and a migration process during which the neuroblasts exit the otic epithelium and aggregate in the periotic mesenchyme. The statoacoustic ganglion complex is the precursor of the spiral (auditory) and Scarpa's (vestibular) ganglia, which will send axonal projections into the areas of the developing auditory and vestibular sensory receptors peripherally and into the brain nuclei of the auditory and vestibular pathways centrally. There are two types of afferent neurons found within the spiral ganglion (discussed below). The top (dorsal) half of the developing otocyst is known as the pars superior, and it produces the sensory receptor structures of the vestibule: three cristae [anterior (superior), posterior (inferior), horizontal (lateral)] and one macula (utricule). The morphogenesis of the three semicircular ducts and the utricle begins during the seventh week of gestation with two diverticuli forming from the dorsal and the lateromedial aspects of the pars superior. These diverticuli elongate and are then pressed together at their centers by proliferating mesenchyme to form areas of fusion. After the elimination of the epithelial cells from the central areas of these semicircular duct diverticuli fusion plates, the dorsal diverticulum forms the anterior and posterior semicircular ducts with a shared crus communes, and the lateral plate forms the horizontal semicircular duct. The ampullary swellings of all three semicircular ducts develop at one end of each of the ducts as they enter the utricle, and a crista ampullaris develops within each of the ampullae. The three cristae and the two maculae are populated with two different types of sensory hair cells: type 1 hair cells bear calyx-type afferent nerve endings, which envelop several hair cell somas, and efferent nerve endings synapse on the calyx and not directly on hair cell somas; the type 2 hair cells bear button-type afferent and efferent nerve endings, which both synapse directly on hair cell somas. The hair cells possess both a single kinocilium and a bundle of stereocilia, composed of modified microvilli, not true cilia. The stereocilia are in graduated height rows, with the tallest rows located closest to the kinocilium. The bottom (ventral) half of the otocyst, known as the pars inferior, produces the scala media of the cochlea organ of Corti and the saccule with its macula epithelium. The saccule begins its development during the seventh week of gestation from the most dorsal aspect of the pars inferior. Development of the cochlea's duct begins in the fifth week of gestation, and its morphogenesis to a full 2¾-turn coiled cochlea is complete by the end of the eighth week of development. The cochlear duct grows out from a proliferative site where it joins the pars superior, so that the oldest sensory cells in this duct are found at its apex and the youngest at its base. The organ of Corti located within the scala media is innervated by the afferent neurons of the spiral ganglion, where ~90% of the neurons are type 1 and the remaining 10% are type 2 neurons. There are two types of sensory hair cells in the organ of Corti: the inner hair cells (a single row), which are innervated by the type 1 spiral ganglion neurons (i.e., 15 to 20 type 1 neurons innervate a single inner hair cell), and the outer hair cells (3 rows), where the pattern of innervation is predominantly efferent with only sparse innervation by afferent fibers of the type 2 spiral ganglion neurons (i.e., one neuron innervates five to 15 outer hair cells). The efferent fibers originate from the lateral and medial efferent systems and innervate the inner and outer hair cells, respectively. The developing auditory hair cells possess both a true kinocilium and polarized arrangements of short stereocilia, but as these sensory cells mature, the kinocilium atrophies; in an adult auditory hair cell, there is only a basal body left to mark the position where the kinocilium had been within the hair cell. The pattern of stereocilia on the immature surface of the hair cells is not organized, but as the hair cells mature, the stereocilia form a straight-line "pipe organ"

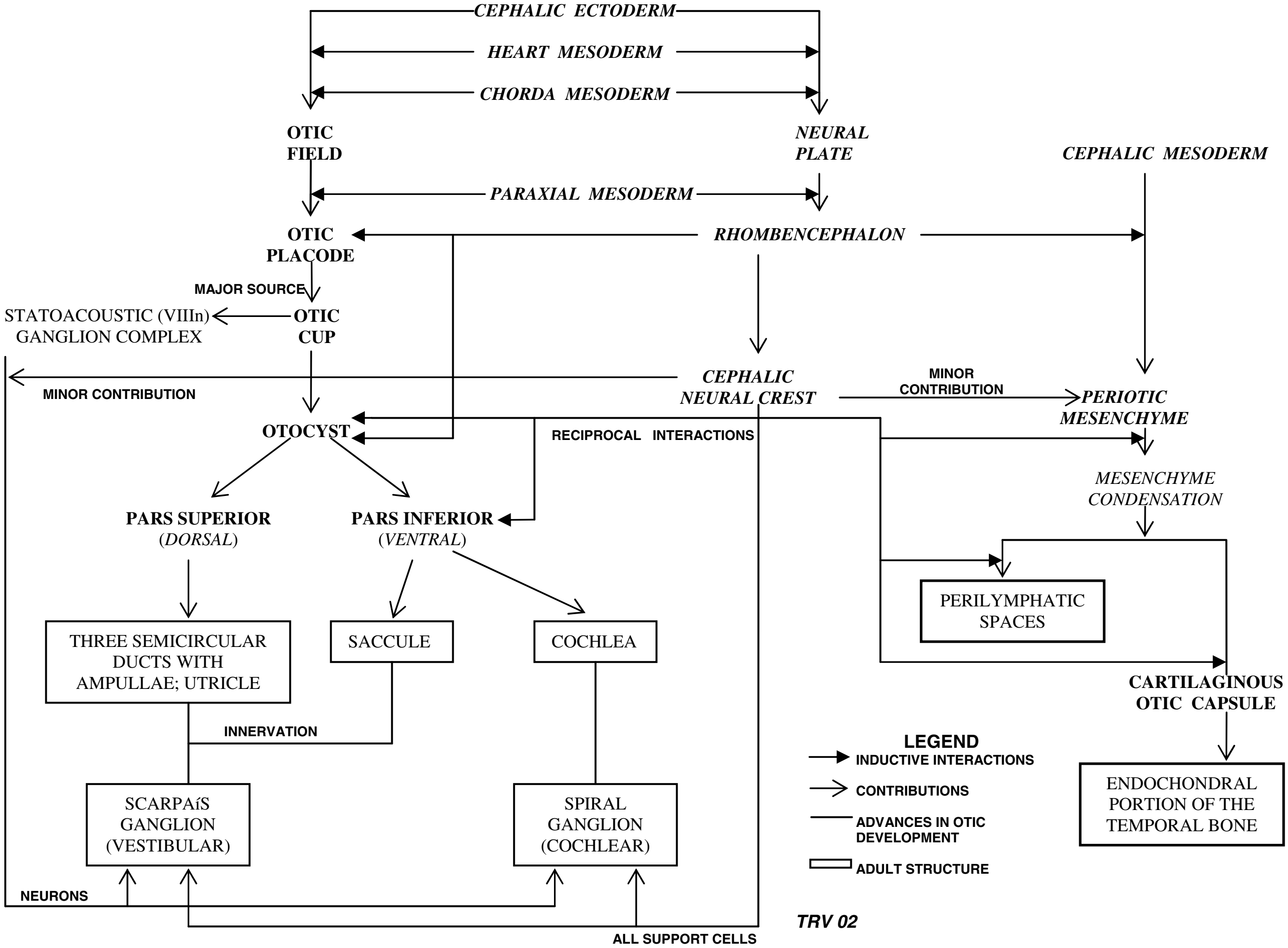


Figure 20–5 *A schematic of the tissue interactions and tissue contributions that occur during the formation of the components of the membranous and bony labyrinths of the mammalian inner ear.*

arrangement on the inner hair cells, and the rows of graduated height stereocilia assume either a "W" or a broadly curved "U" pattern on the cuticular plates of the outer hair cells. During the fifth week of gestation a diverticulum that originates from the top of the dorsal half of the otocyst elongates and forms the precursor of the endolymphatic duct and sac. The fluid within the membranous labyrinth is filled with endolymph (high potassium, low sodium). The maintenance of the unique ion concentration of endolymph is thought to depend on the proper functioning of the stria vascularis. The stria vascularis is composed of marginal cells that border the scala media lumen and have their origin from the ectodermal cells of the otocyst; intermediate cells, which are the next layer of internal cells, originate from cephalic periotic mesenchyme. The basal cells of the stria, which interface with the spiral ligament, also originate from the periotic cephalic mesenchyme. The neural crest contributes melanocytes to the stria vascularis during its formation, and these melanocytes are thought to be important to the inner ear's ability to resist damage by oxidative stress (e.g., aminoglycoside ototoxicity). The stria vascularis is anchored to the bony wall of the temporal bone by the spiral ligament. The spiral ligament originates from cephalic mesenchyme, has a high level of metabolic activity, and is known to express many important growth factor and antioxidant genes, and its fibrocytes (types I–IV) are thought to contribute to ion recycling from the auditory hair cells within the scala media. The membranous labyrinth is protected by a fluid space filled with perilymph (low potassium, high sodium), and this space develops from the periotic head mesoderm that surrounds the developing otocyst. The perilymphatic spaces and the cartilaginous otic capsule, which is the anlage of the petrous portion of the temporal bone, are both the products of the inductive tissue interactions between the otic epithelium (of the developing otocyst) and its surrounding periotic mesenchyme, which occurs during the ninth week of development. At the same time that the otocyst is inducing mesodermal aggregation to form a cartilaginous otic capsule it is also turning off the chondrogenic phenotype in a subset of the aggregating mesenchyme that is closest to the otocyst, so that this area can be cleared of these mesenchymal cells (by either programmed cell death or cell migration) to form the perilymphatic spaces of the labyrinth. This occurs during the 12th through 16th weeks of gestation. The lining cells of the perilymphatic cavities are derived from the surrounding periotic mesenchyme. During the ninth week of development the cartilaginous otic capsule is formed from the surrounding periotic mesenchyme. The endochondral bone of the petrous portion of the temporal bone is the result of an osteogenic replacement process, in which cartilage is replaced with endochondral bone, as evidenced by the presence of isolated cartilaginous rests within the petrous portion of the temporal bone. The internal auditory canal is formed at the stage of otic capsule chondrogenesis when the seventh and eighth cranial nerves interact with the nearby population of chondrocytes (much like the otic epithelium interacts with these cells) to turn off their chondrogenic program to specify a cell-free channel for these nerves. The mastoid and squamous portions of the temporal bone are formed by the laying down of membranous bone.

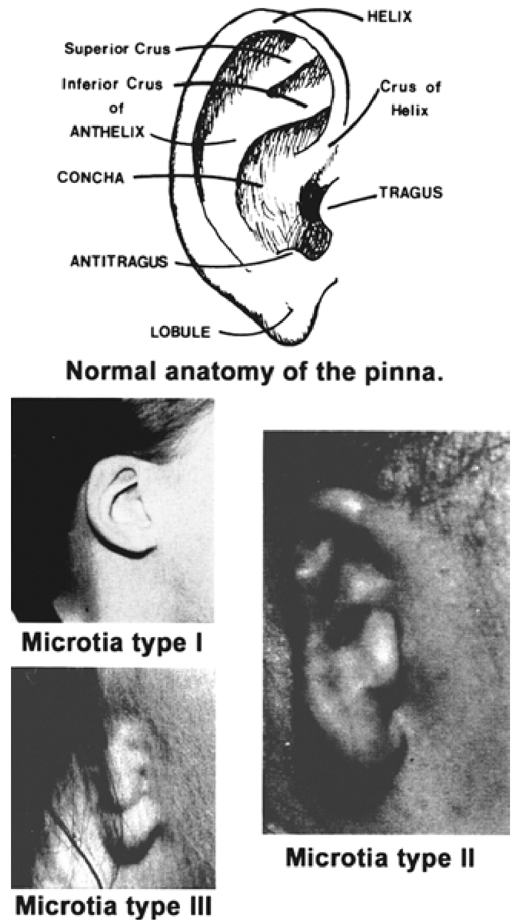


Figure 20–6 *A schematic of the anatomy of a normal pinna (external ear) and a series of macrophotographs depicting types I, II, and III microtia malformations of the pinna.*

CONGENITAL MALFORMATIONS

External Ear

Pinna

The most common malformation of the external ear is microtia. Abnormalities of the pinna span the spectrum from diminutive pinnae with an essentially normal configuration (type I) to tags of skin that have no recognizable features and are accompanied by atresia of the external auditory canal (type III) (**Fig. 20–6**). There is a relationship between pinna malformations and the malformation of other first and second branchial arch derivatives (i.e., middle ear ossicles). However, it is important to remember that a severe malformation of the pinna with accompanying atresia of the external auditory canal is not always accompanied by a disruption in the development of an intact ossicular chain, the middle ear cavity, and the inner ear. Even though segments of the pinna and the ossicular chain both have there origins from the same branchial arches, they are from different segments of neural crest cells and emerge at different times in the development of the embryo. As a rule of thumb, the more severe the malformation of the pinna, the more likely it is that there are associated problems with the development of the components of the middle ear and the inner ear. However, it is important not to assume anything and to obtain additional tests that define the integrity of the ossicular chain and middle ear cavity and the functional capacity of the inner ear.

Certain types of pinna malformations are associated with specific disorders or syndromes: for example, a round hypoplastic pinna is indicative of Down syndrome (**Fig. 20–7A**); often the pinna of infants with trisomy 13–15 has a sharp posterior angulation of the crus helix with hypoplasia of the antihelix and lobule (**Fig. 20–7B**). A "railroad track" pinna is a subtle malformation that is characterized by a marked prominence of the crus helix forming a diagonal bar across the concha, parallel to and as high as the antihelix, and is often present in infants suffering from fetal

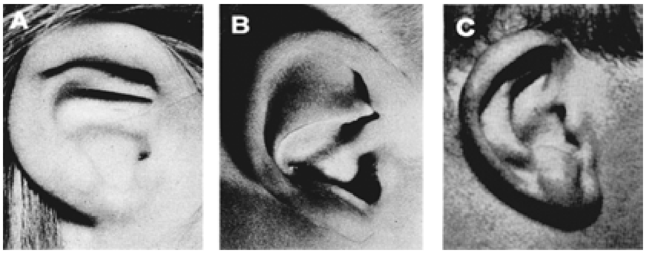


Figure 20–7 *Macrophotographs depicting the phenotypic appearance of pinnas from **(A)** a Down syndrome child, **(B)** a fetal alcohol syndrome child, and **(C)** a child with a severe craniofacial malformation (i.e., Treacher Collins syndrome).*

alcohol syndrome (i.e., 30%). Excessive thickening of the lobule can be an indicator of malformations that involve the long process of the incus and the head of the stapes; absence of the superior crus of the helix can be indicative of congenital fixation of the ossicular chain (**Fig. 20–7C**). Either severe malformation of the pinna or anotia is associated with various syndromes that can cause severe craniofacial malformations (e.g., Treacher Collins and Alport's syndromes). In sum, it is important to note the form and the level of normal development achieved by the pinna because it can direct the evaluation of the pediatric patient; however, it is equally important to appreciate that normal development of the middle and inner ears can occur even when there is a severe malformation of the pinna of the external ear.

External Auditory Canal

Atresia of the external auditory canal may be either osseous or membranous and either partial or complete. Atresias, either partial or complete, are correctable, but it is important to determine the condition of both the ossicular chain and the middle ear cavity before contemplating a surgical repair. Another congenital anomaly of the external auditory canal that can impair hearing is stenosis of this canal, in which the lumen is greatly reduced in circumference. This condition carries with it a greatly increased risk of formation of cholesteatoma and canal stenosis can go undetected because it does not always involve a malformed pinna to alert the physician.

Middle Ear

Middle Ear Cavity

Both congenital cholesteatomas and osteomas can cause hearing loss in the affected individual. Congenital cholesteatoma can originate from either rests of embryonic cells left in the middle ear cavity or leftover cellular elements from amniotic fluid. An osteoma is a benign bony growth that results from aberrations of bone formation and remodeling that occur during the development of an infant's middle ear cavity. A conductive hearing loss will result from the presence of either of these masses in the middle ear.

Ossicles

An absence of either all or some of the ossicles has been associated with syndromes such as achondroplasia and diastrophic dwarfism. Fracturing of the ossicles can occur in the pediatric population with osteogenesis imperfecta (van der Hoeve's syndrome). Congenital otosclerosis can result in fixation of the stapes and an accompanying absence of the oval window; this is primarily a disorder of the mesodermal mesenchyme of the otic capsule. Developmentally, the stapedial artery passes through the foramen of the stapes and normally atrophies with maturation of the middle ear. If persistent, it can interfere with the conduction of sound. X-linked disorder of the stapes has an associated problem of perilymphatic gusher that can occur during inner ear surgery. In all craniofacial disorders that involve the first and second branchial arches (e.g., mandibulofacial dystosis), it is prudent to assume that there may be some effect on the integrity of the ossicular chain, and the examining physician should always test for a conductive hearing loss.

Inner Ear

Developmental anomalies of the inner ear are usually combined with other malformations and frequently have a genetic component. Inner ear malformations are usually classified with regard to their severity. The following is a well-recognized system of classification and is presented in an order of decreasing severity: (I) Michel's anomaly indicates the complete lack of any labyrinthine form; (II) Mondini's anomaly affects both the membranous and the bony labyrinths, resulting in an immature vestibular labyrinth and a hypoplastic cochlear duct that has a reduced number of coils; (III) Bing-Seibenmann's anomaly has a hypoplastic membranous labyrinth but a normal-appearing bony labyrinth; and (IV) Scheibe's anomaly refers to the hypoplastic development of the cochlea and saccule, which derive from the pars inferior (ventral) portion of the otocyst, but a normal-appearing vestibule, which originates from the pars superior (dorsal) portion of the otocyst. Some specific effects of chromosomal anomalies on the development of the inner ear are (1) trisomy 13, which can cause hypoplasia of both the saccule and the cochlea, and (2) trisomy 18, in which there is an absence of the spiral ganglion. Acquired anomalies of the inner ear include both the action of known teratogens such as retinoic acid and thalidomide and the sequelae of prenatal infections. Several examples of anomalies that result from prenatal infections include rubella, which is known to cause Scheibe-type anomalies; cytomegalovirus, which can cause damage to the vasculature of the cochlea and produces inclusion bodies in the cells of both the utricle and saccule; and toxoplasmosis, which can cause calcification of the spiral ligament of the cochlear duct.

Suggested Readings

Schutneckt HF. Dysmorphogenesis of the inner ear. In: Gorlin RJ, ed. Morphogenesis and Malformation of the Ear (March of Dimes Birth Defects Original Article Series, Vol. 16, No. 4). New York: Alan R. Liss; 1980:47–71

Van De Water TR, Noden DM, Maderson PFA. Embryology of the ear: outer, middle. In: Alberti PW, Ruben RJ, eds. Otologic Medicine and Surgery. Vol 1. New York: Churchill Livingstone; 1988:3–27

Warkany J. Aural defects. In: Congenital Malformations: Notes and Comments. Chicago: Year Book Medical Publishers; 1971: 401–416

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. The dorsal (pars superior) portion of the otocyst forms
   A. The cochlea
   B. The bony portion of the labyrinth
   C. The semicircular ducts and the utricle
   D. The ossicular chain
   E. The tympanic membrane
2. The first pharyngeal pouch forms
   A. The semicircular ducts
   B. The ossicular chain
   C. The cochlea
   D. The saccule
   E. The middle ear cavity
3. A malformed pinna is
   A. Always associated with middle ear anomalies
   B. Pathognomonic for an atrophic cochlea
   C. An indication of first and second branchial arch abnormalities and may be associated with abnormalities of the ossicles
   D. A sign of severe anomalies of the central auditory pathway
   E. Never associated with abnormalities of the middle ear and the ossicles
4. Scheibe's anomaly is best described as
   A. Agenesis of the inner ear
   B. Malformation of the vestibule
   C. A congenital malformation of the stapes
   D. Dysplasia of the cochlea and the saccule
   E. Stenosis of the external auditory canal
5. Rubella infection of the inner ear can result in
   A. Scheibe-type dysplasia of the cochlea and saccule
   B. Agenesis of the stria vascularis
   C. Malformation of the anterior crista
   D. Agenesis of a cochlear duct
   E. Duplication of the lateral semicircular duct

Chapter 21

ACOUSTICS AND MIDDLE EAR MECHANICS FOR OTOLARYNGOLOGY

John J. Rosowski and Saumil N. Merchant

The physical properties that define how sound propagates in air are of great consequence in how we perceive sound. This chapter will review some of the elementary physical processes associated with sound propagation in air and the reception of sound by human ears. The chapter is not meant to provide indepth details about physical processes of sound and its reception but is instead meant to impart some very basic perceptual consequences of how sound propagates in air, how sound is collected by the ear, and how these processes are affected by pathology. Readers interested in a more detailed description of sound and the ear could read some elementary texts on the subject (see suggested readings, Yost, 1994).

WHAT IS SOUND?

Sound results when particles of a medium are set into vibration. The sounds we hear, for the most part, result from air particles being set in motion by vibrating objects. For example, the vibrating tines of a struck 512 Hz tuning fork (**Fig. 21–1**) produce backward and forward motions of the air particles that surround the tines. (An air particle is a small volume of air that contains many air molecules.) The particles are set in motion by the vibrating tines, then push on the air particles next to them, where the push is proportional to the sound pressure, setting the next layer of particles into back-and-forth motion. The physical disturbance of sound pressure and particle motion, not the particles themselves, propagates through the air medium as succeeding layers of air particles are set into vibration. The amplitude of the propagating physical disturbance can be quantified in terms of either the sound pressure acting on the particles or the amplitude of the particle motions. The dependence of sound propagation on the presence of a medium explains why sound does not propagate through a vacuum.

A 512 Hz tone that is just audible to many humans will be associated with pressure variations of ~0.2 millipascal (a pascal is a unit of pressure, see later) and back-and-forth particle displacements of ~16 angstroms (Å). In general, the louder the sound, the larger the particle motions and pressure variations. In practice it is easy to measure the pressure variations and more difficult to measure the motion of air particles, so sound pressure is the primary measure of sound amplitude.

What Is Sound Pressure?

Sound pressure refers to the magnitude of the cyclic variations in pressure produced around ambient static pressure when a sound is produced (**Fig. 21–2**). A pressure has units of force per area. The international unit of pressure is the pascal (Pa), where 1 Pa = 1 newton (N) of force per square meter of area. A newton is the force necessary to accelerate a kilogram 1 m/s^2. This is a moderate force, but the pressure produced by spreading a newton over a square meter of area is very small. In fact, 1 Pa of pressure equals only 1×10^{-5} atmospheres. The quietest sounds we hear are of even lower pressure; the change in pressure associated with sound at the threshold of hearing for a 1000 Hz tone is ~20 micropascals (μPa) or 2×10^{-10} atmospheres.

There are several ways to quantify sound pressure, though the most common is in terms of the root mean square (rms) deviation in pressure. **Fig. 21–2A** is an illustration of the changes in pressure with time that are associated with a 512 Hz tone. This figure illustrates the simple relationship between peak, peak-to-peak, and rms measures of amplitude of a pure tone where these different measures of amplitude yield values of 1, 2, or 0.71 Pa,

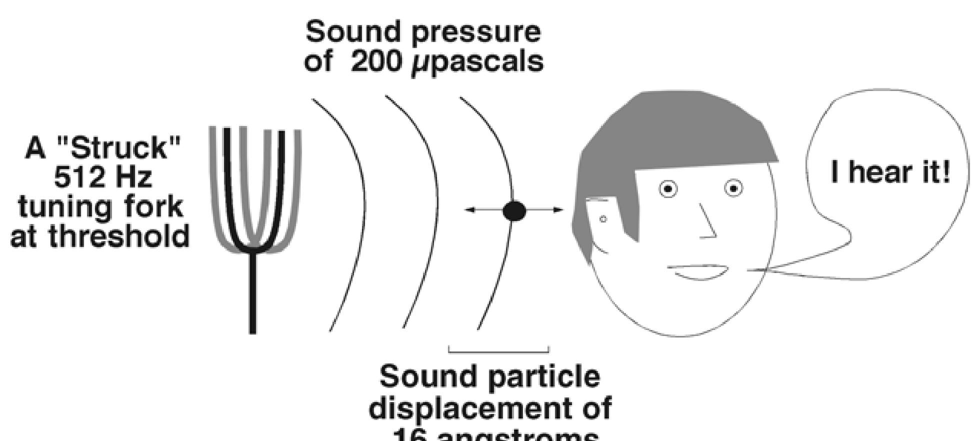


Figure 21–1 *A struck tuning fork sets nearby air particles into motion with a frequency equal to the natural frequency of the fork. Associated with the motion of the air particles is a sound pressure. The air particles that are set in motion push on the particles next to them, and so forth, resulting in a propagating physical disturbance. A tone of 512 Hz that is just audible is associated with air particles in free space moving back and forth with a magnitude of 16 angstroms (Å) (a displacement that is a little larger than the diameter of a hydrogen atom) and pressure oscillations of 200 micropascals (μPa)*

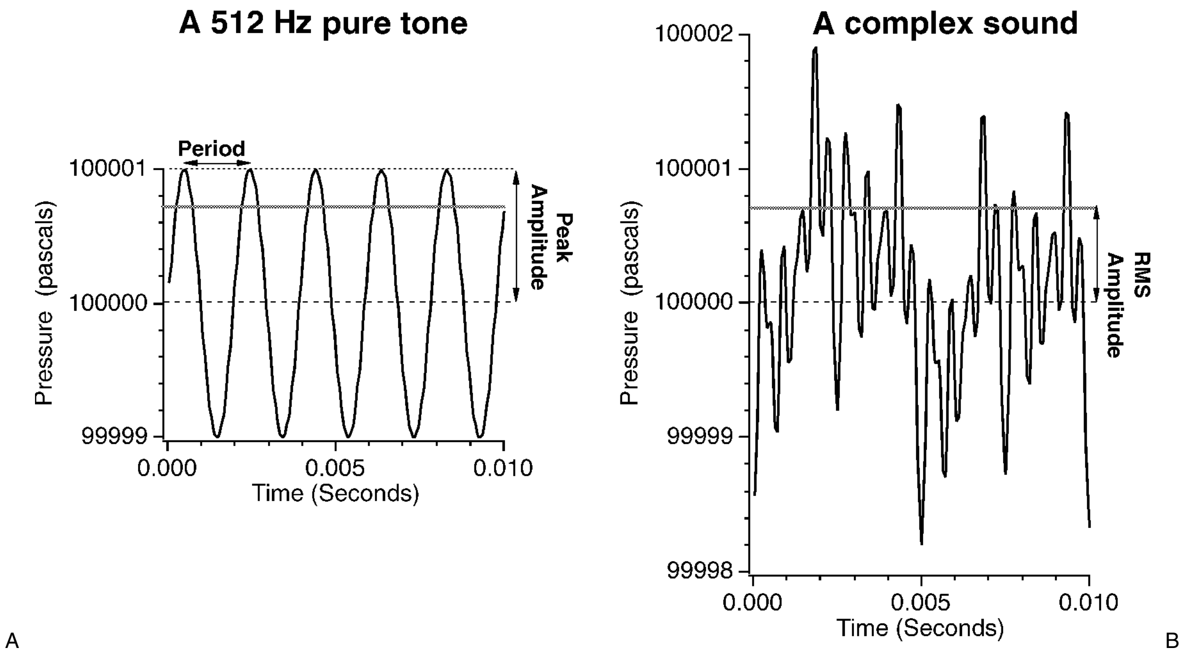


Figure 21–2 *Two patterns of temporal variations in pressure in air produced by sound. The absolute value of the pressure is scaled in both plots, and therefore the sound-induced variations occur around a static value of 100,000 Pa = 1 atmosphere. The sound pressure corresponds to the amplitude of the variations around the static value; in the illustrated cases, while the static atmospheric pressure is 100,000 Pa, the amplitude of the sound pressure is on the order of 1 Pa.* ***(A)*** *The pressure variations are those of a 512 Hz tone. The pressure varies sinusoidally with a period of 1/512 = 0.00195 sec. The amplitude of the pressure variation about the static value can be quantified in terms of: the peak-to-peak value of 2 Pa, the peak value of 1 Pa, or the root mean square (rms) value of 0.707 Pa rms. (The root mean square value of the variation in pressure is the square root of the mean of the squared pressure deviations from static values averaged over some time. In the case of a sinusoid, a convenient averaging time is some integral number of periods of the sine wave. With sinusoidal sound pressures, the rms value equals the peak amplitude/$\sqrt{2}$.)* ***(B)*** *The pressure variations are those of a complex sound with many irregular risings and fallings of the sound pressure. With this kind of sound, peak amplitude and peak-to-peak amplitude are poor indicators of the average sound level. However, rms is an excellent measure as long as one specifies an averaging time; in this case, we compute the rms sound pressure over the 0.01 sec time window shown in the illustration. Note that the sound pressure in* ***B*** *has the same rms value as the sound pressure in* ***A****.*

respectively. The time-varying sound pressure waveform illustrated in **Fig. 21–2B** is not readily quantified by peak measures but is well described by its rms value. In fact, the rms sound pressure of this complex sound is 0.71 Pa, a value identical to the tonal sound pressure observed in **Fig. 21–2A**.

The human auditory system is sensitive to a wide range of sound pressures: The minimum sound level that a human can hear is between 10 and 20 rms μPa. Conversational speech has pressures that are ~100 times threshold. A lecturer or loud speaker will talk with sound pressures 1000 times threshold. Music often contains sound pressures that are 10,000 times threshold. Guns, fireworks, and jet engines can produce pressures that are more than 1,000,000 times threshold. Because of this sensitivity to pressures that vary by more than 1,000,000 times, and because our ear is generally sensitive to fractional changes in pressure, we commonly use a logarithmic scale to grade pressures. The decibel (dB) is a logarithmic measure of relative energy, where 10 dB (1 bell) represents an increase over a given reference energy level of one order of magnitude (i.e., one common log unit or a factor of 10). The common reference level for sound pressure level (SPL) is 2×10^{-5} Pa, and because energy is proportional to pressure squared:

$$\text{Sound level in dB SPL} = 10 \log_{10} \left(\frac{X}{0.00002 \text{ rms Pa}}\right)^2$$

$$= 20 \log_{10} \left(\frac{X}{0.00002 \text{ rms Pa}}\right),$$

where X is the sound pressure in rms pascals. Different dB sound pressure scales use different reference pressures [e.g., the dB hearing level scale (dB HL) of the

TABLE 21–1 THE SOUND PRESSURES OF COMMON SOUNDS

Approximate Sound Level		
rms Pa	**dB SPL**	**Sound Source**
0.0001–0.0002	14–20	Just audible whispers
0.002–0.02	40–60	Conversational speech
0.02–0.6	60–90	Noisy room
0.6–20	90–120	Loud music
>20	>120	Gunfire

dB, decibel; Pa, pascal; rms, root mean square; SPL, sound pressure level.

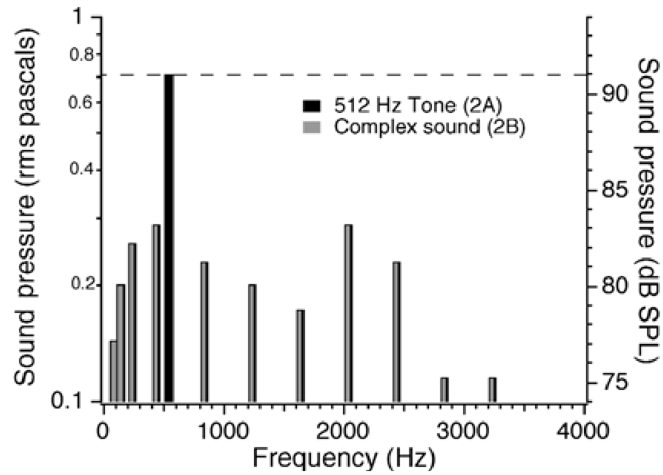


Figure 21–3 The Fourier spectra of the two sound signals from ***Fig. 21–2*** *are illustrated. The black line shows the line spectral magnitude at 512 Hz associated with the 512 Hz pure tone of* ***Fig. 21–2A****. The gray lines show the magnitudes of the spectral components of the complex sound of* ***Fig. 21–2B****. The linearly scaled abscissa or horizontal axis describes the frequency of the different components. The logarithmically scaled left-hand vertical axis or ordinate describes the rms magnitude of the different spectral pressure components. The linear dB scale on the right calibrates the pressure magnitude of the components in terms of dB SPL.*

clinical audiogram uses the average sound pressure threshold of the normal population at a given frequency as the reference pressure]. The SPL of both of the sounds in **Fig. 21–2** is 91 dB SPL, where 91 = 20 log10 (0.71/0.00002). The sound pressures of various commonly experienced sounds are noted in terms of rms pascals and dB SPL in **Table 21–1**.

WHAT IS SOUND FREQUENCY?

Sound varies not only in amplitude but with time. One of the simpler sounds is a pure tone in which the relationship between time and sound pressure is described by a sine function, as in **Fig. 21–2A**, where the sound pressure, the variation in pressure about atmospheric pressure, equals $\sin(2\pi 512t)$, where t is time in seconds. The human ear is sensitive to sound frequencies of 20 to 20,000 Hz. Complex sounds (e.g., **Fig. 21–2B**) can be described by the addition of pure tones of different frequencies and different phases. One can investigate the frequency content of sounds by using a Fourier spectrum analyzer that converts temporal responses into frequency spectra. **Fig. 21–3** shows the spectra of the magnitudes of the two time waveforms of **Fig. 21–2**. The 512 Hz tone of **Fig. 21–2A** gives rise to a line spectrum at 512 Hz with no other spectral components. The magnitude of the sound pressure spectral component of the tone is 0.71 rms pascal (the left-handed scale of **Fig. 21–3**) or 91 dB SPL (scaled on the right). The superimposed spectral analysis of the complex sound of **Fig. 21–2B** demonstrates that the sound is made up of multiple components with varied pressure magnitude.

It should be noted that although the spectra of **Fig. 21–3** describe the magnitudes and frequencies of the different spectral components that make up the sounds of **Fig. 21–2**, the illustrated spectra do not contain enough information to reconstruct the waveforms of **Fig. 21–2**. What is missing is phase information. The phase of a sinusoid describes the time when the pressure crosses from negative to positive. This timing information is critical when we combine waves, because two waves added together can sum constructively if they have similar phases, sum to near 0 if the waves are out of phase and of identical magnitude, or anything in between. Many spectral analyzers also yield phase information.

TO WHAT SOUNDS ARE OUR EARS SENSITIVE?

Every human ear is differentially sensitive to sounds of different frequency, but measurements of human sensitivity to sound vary depending on how we specify the sound stimulus. **Fig. 21–4** compares two different measurements of the mean sound pressure threshold measured in normal young adults with pure tones of different frequencies. (The sound pressure threshold is the lowest sound pressure that is audible.) The lower curve in **Fig. 21–4** depicts sound pressure measurements made with the listeners in an anechoic room where the sound pressure measurement was made where the listener's head would be when the listener was not present (a "free field" description of the sound

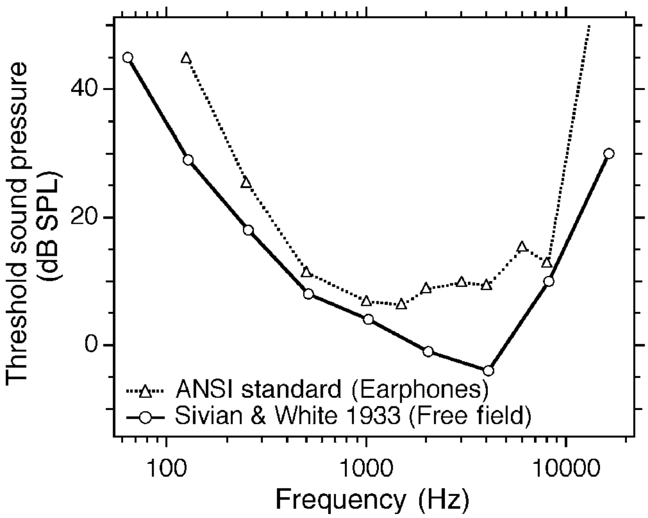


Figure 21–4 The sensitivity of the ear to sounds of different frequencies. This figure is a comparison of two studies of auditory threshold, those combinations of sound pressure level and frequency that are just audible: The American National Standards Institute (ANSI) description (1969) of normal hearing threshold made under earphones and the free-field studies of Sivian and White (1933) made in an anechoic room. The sound pressure scale on the ordinate is in units of dB SPL. Frequency, on the abscissa, is plotted on a log scale. The mean normal threshold at 1000 Hz in both measurement conditions is ~0 dB SPL.

pressure). The upper curve is the American National Standards Institute (ANSI) standard measurement of human thresholds made under earphones, where the sound pressures are those generated by the earphones in a calibration coupler. The differences between these two thresholds can be explained by the effect of the human listener on the open sound field, the effect of closing the ear canal by earphones, and differences in calibration between the two circumstance (Shaw, 1974).

Whether under earphones or in the "free field," however, it is clear that normal young adult humans are most sensitive to sound frequencies of 500 to 8000 Hz. The best frequency differs depending on the measurement circumstance, being equal to 1500 Hz under earphones and 4000 Hz in the free field. At higher and lower frequencies, more sound pressure is required to induce a threshold response, and the thresholds move up steeply below 500 Hz and above 8000 Hz.

Otologists and audiologists are usually most interested in how measurements of an individual's hearing thresholds differ from normal, where under earphones normal is practically defined by the ANSI standard measurements of **Fig. 21–4**. The most powerful graphical tool for comparing two different functions is to plot their difference. The clinical audiogram (**Fig. 21–5**) uses this difference technique by plotting the individually determined sound threshold relative to the ANSI normal hearing level. For example, a person who is shown to have a hearing threshold at 1000 Hz of 10 dB greater than the ANSI standard would be assigned a hearing level of 10 dB at that frequency. In clinical audiograms, sound pressure relative to the standard is plotted on a decreasing scale in dB relative to normal, and frequency is tested in octave or half-octave intervals. Remember that the normal curve is based on mean thresholds and that there is normal variation about the mean.

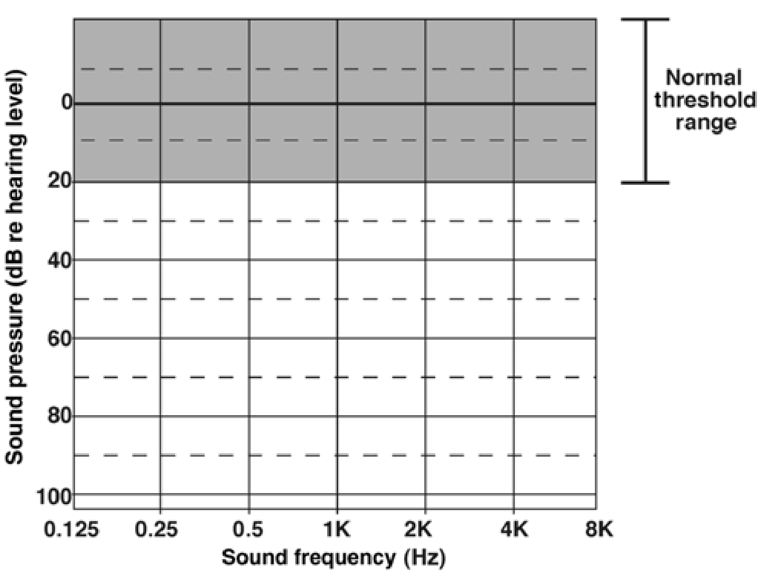


Figure 21–5 The clinical audiogram. The sound pressure scale is relative to the ANSI standard description of normal human hearing thresholds, and the scale is inverted so that higher thresholds are plotted lower on the graph. The gray-shaded region shows the range of variations in normal listeners.

INFLUENCES OF THE PROPAGATION OF SOUND ON HUMAN PERCEPTION

THE SPEED OF SOUND *c*

The sound-induced vibration of particles propagates through space by pushing on other particles and setting them into vibration. The velocity of propagation of sound through a medium is proportional to the root of the stiffness of the medium (the pressure necessary to change the volume by a set amount) and inversely proportional to the root of the density (mass per unit volume) of the medium. **Table 21–2** compares the stiffness, density, and speed of sound in air, water, and hydrogen. The gases have a similar low stiffness compared with water, and the density varies greatly among the three media. Comparison of the entries in **Table 21–2** argues against the common misconception that sound travels faster in water because water is more dense. In fact, sound travels faster in water because water is much stiffer than air; the increased density of water actually serves to decrease the growth of propagation velocity in water that results from the greater difference in stiffness between air and water. The inverse relationship between density and speed of sound is best observed by comparing sound speed in air and hydrogen. Both hydrogen and air are of similar stiffness, but because hydrogen is less dense, sound travels faster in hydrogen.

The stiffness and density of air depend both on the temperature and static pressure. At 20°C (68°F) and 1 atmosphere of pressure, the propagation velocity for sound in air is ~340 milliseconds (msec), roughly 1 foot per msec. Because the human head has dimensions of ~8 to 10 inches in diameter, there can be appreciable time delay (0.5–1 msec) involved in the propagation of sound from one ear to the other. This delay has a perceptual impact, in that the difference in the time of arrival of sound at the two ears is one of the cues used in judgments of sound source localization.

TABLE 21–2 MATERIAL PROPERTIES AND THE SPEED OF SOUND IN DIFFERENT MEDIA MEASURED AT 1 ATMOSPHERE OF PRESSURE AND 20 DEGREES CENTIGRADE

Material	Density (kg/m^3)	Stiffness (Pa)	Propagation Velocity (m/sec) (speed of sound) $\sqrt{\text{stiffness/density}}$
Air	1.2	1.4×10^5	343
Water	1000	2.2×10^9	1,481
Hydrogen	0.09	1.4×10^5	1,270

SOUND WAVELENGTH AND THE DIFFRACTION AND SCATTERING OF SOUND

The speed or propagation velocity of sound helps define the sound wavelength, the distance it takes a propagating disturbance to repeat itself. In the case of pure tones, the wavelength (usually identified with the symbol λ) is defined by

$$\lambda = \frac{\text{propagation velocity}}{\text{sound frequency}}$$

The wavelength is an important variable in determining how sound interacts with objects (**Fig. 21–6**). If the wavelength of sound is at least five times larger then the

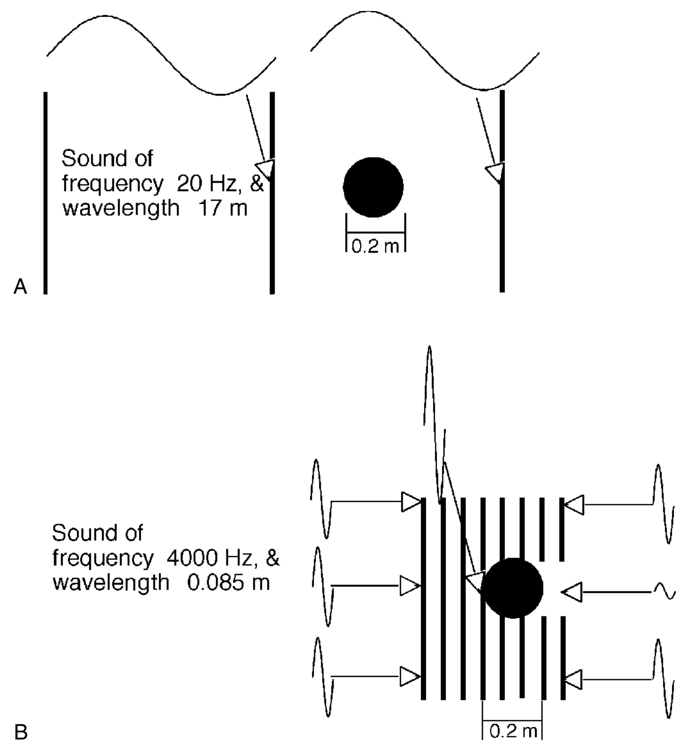


Figure 21–6 *Two schematics illustrating the dependence of significant sound diffraction and scattering effects on sound wavelength. Both schematics depict the interaction of a sound wave propagating left to right with a 20 cm sphere. The propagating sound wave is depicted by the regular vertical bars that denote wave fronts where the sound pressure is 0 at some time t. The waves depict the sound pressure amplitude made at the locations described by the arrows.* ***(A)*** *The wavelength is large compared with the sphere's diameter, and the presence of the sphere has little effect on the amplitude of the propagating wave.* ***(B)*** *The wavelength is small compared with the diameter of the sphere. In this case, the interaction of the wave with the sphere causes an increase in the sound pressure on the forward surface of the sphere and a sound shadow area on the back surface where the sound pressure is smaller than the pressure in the propagating wave.*

largest dimension of an object, the object will have little effect on the sound; that is, as the sound propagates around the object, the sound pressure at the front and back side of the object will be very similar to the sound pressure measured when the object is not present. In contrast, if the wavelength is similar or smaller than the dimensions of an object, that object will introduce variations in sound pressure into the sound field. In general as the short-wavelength sound interacts with the object, the sound pressure along the front surface of the object, where the sound first falls incident, will increase because of the reflection of sound backward, whereas the sound along the back surface will be decreased. A common analogy is between light and sound, where in the small wavelength case, the object casts a sound shadow.

The size of our body structures relative to the sound wavelength plays a significant role in determining how we and our ears interact with sounds of different frequencies (Shaw, 1974). A 20 Hz sound wave with a wavelength of 17 m is little affected by our head or body. The sound in a 200 Hz wave with wavelength of 1.7 m can be effectively scattered by the shoulder and torso, with the result a small gain in sound pressure at the ear. The sound in a 2000 Hz wave with a wavelength of 17 cm is diffracted by the head, so that there is a doubling of the sound pressure on the side of the head directed toward the sound source and a shadow on the opposite side of the head (**Fig. 21–6B**). Four thousand Hz tones of 8.5 cm wavelength are scattered by the pinna of the external ear in a manner that increases sound pressure for sound sources pointing directly at the ear opening and decreases the pressure for other directions. Another kind of wavelength interaction is apparent in the external ear; resonances occur within the ear at frequencies where the length of the ear canal and depth of the concha are odd multiples of $\lambda/4$ (Shaw, 1974). **Table 21–3** lists some of the critical frequencies above which the sound wavelengths allow interactions with various parts of our body and ear.

Table 21–3 Wavelengths and anatomic structures with comparable dimensions

Frequency (Hz)	Wavelength	Anatomical Structure	Structural Dimensions
200	1.7 m	Torso	0.5 m
2000	17 cm	Head	10 cm
4000	8.5 cm	Pinna ear canal length	4 cm 2.5 cm
20,000	1.7 cm	Ear canal diameter	0.8 cm
		Tympanic membrane	

Variations of Sound Pressure with Distance from a Source

As opposed to the straight, parallel planar wave fronts seen in **Fig. 21–6**, most sounds propagate as spherical waves, where the pressure amplitude and phase are identical at any radial distance r from the sound source. The magnitude of the pressure (P) in a spherical wave front falls regularly with distance from the sound source, where $P(r)$ is proportional to $1/r$. This inverse dependence on distance from the source describes a fractional dependence in which sound pressure falls by half for every doubling of distance. In a free field, then, in order to be heard, you must speak louder as you move away from a listener. As you get farther away but continue to move, the listener will notice smaller changes in sound pressure because you need to move more to produce a significant fractional change in your distance from the listener. It should be noted that except for outside concerts and large lecture halls, human speakers and listeners are rarely anywhere like the free field. In our houses, classrooms, and conference rooms, sound pressure reverberates off the walls, floor, and ceiling, producing a significant increase in the sound pressures that reach a listener's ears.

THE EAR AS A COLLECTOR OF SOUND

Overview

The external and middle ear are known to have a significant influence on the sensitivity of the human ear to specific sounds. Although it is true that the inner ear is the most important element in determining whether the ear is sensitive to a specific sound frequency and level, the inner ear will not be able to respond with any sensitivity to signals that are greatly attenuated by the middle and external ear. Furthermore, signals that are best collected by the external and middle ear generally correspond to the most sensitive frequencies of hearing in humans and other animals (Rosowski, 1991). This section of the chapter reviews some of the acoustic and mechanical mechanisms used by the external and middle ear to gather sound and conduct it to the inner ear. Again, this section is meant to be an overview rather than an exhaustive or detailed review of this topic. Those readers wishing more detailed descriptions should seek other sources (see suggested readings, Dallos, 1973; Rosowski, 1996; Zwislocki, 1975).

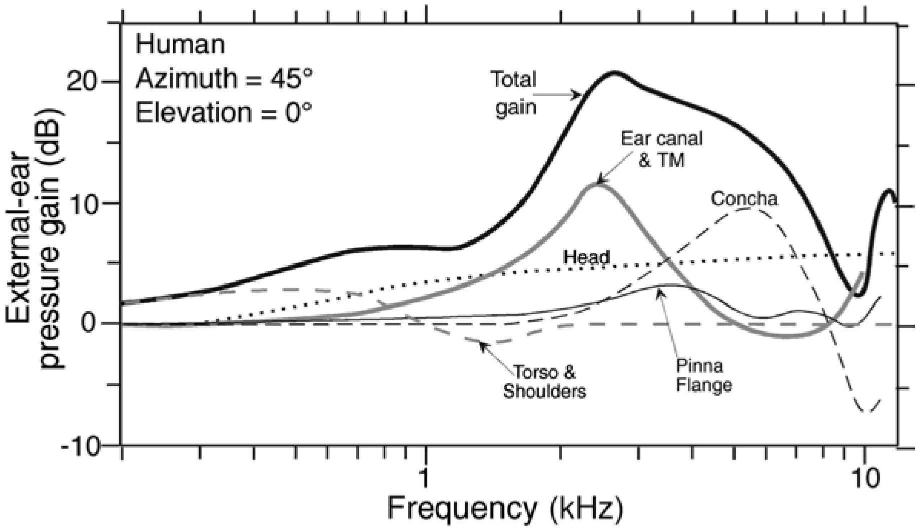


Figure 21–7 A schematic representation of the contribution of different anatomic features to the external-ear gain. The total gain and the gain of the individual components (in dB) is plotted versus frequency. The schematic describes the different gains for a sound source that is positioned on the same horizontal plane as the interaural axis (elevation of 0 degrees) and 45 degrees off the midline (the azimuth) toward the ear that is measured. The gains of the different components are all multiplied (added in dB) together to achieve the total gain. TM, tympanic membrane. (After Shaw, 1974.)

The External Ear

The external ear, as well as the head and body, has a significant influence on the sounds that reach the middle ear. This acoustic function of the external ear can be described by a frequency and directionally dependent alteration in the sound pressure at the tympanic membrane when compared with the sound pressure in the free field, sometimes called the external-ear gain. **Fig. 21–7** illustrates that for sound sources that are positioned facing the ear, the external ear produces a gain in sound pressure of as much of 20 dB at 4000 Hz, with less gain at lower and higher frequencies. **Fig. 21–7** also illustrates that this gain results from the combination of sound scattering and diffraction around the head and torso, as well as the acoustic influence of the pinna, concha, and ear canal tube. Not shown in the figure is how this external-ear gain is directionally dependent especially above 1000 Hz. In fact, for sounds coming from the opposite side of the head, there is little overall external-ear gain.

The Tympano-ossicular System

Another source of sound pressure gain is the impedance–transformation that is associated with the tympano-ossicular system of the middle ear. **Fig. 21–8** schematizes the key features in this transformation, where the difference between the acoustic properties of air and cochlear fluid are coupled together by the tympanic membrane and ossicles. The key transformer within the middle ear is the ratio of the tympanic membrane and footplate area. The tympanic membrane gathers force over its entire surface and then couples the gathered force to the smaller footplate of the stapes. Because pressure is force per area and the human tympanic membrane has an area that is 20 times larger than the footplate, the sound pressure applied to the inner ear by the stapes footplate can be as much as 20 times (26 dB) larger than the sound pressure collected by the tympanic membrane. The lever action of the different lengths of the rotating malleus and incus arm also has the potential for adding some small pressure gain. (The malleus and incus lever arms in humans are nearly the same length, and the ossicular lever is estimated to add only 2 dB of ossicular gain.)

Measurements of the middle ear pressure gain of the human middle ear are illustrated in **Fig. 21–9**. These data point out that the real middle ear generally produces less gain than that predicted by the anatomical transformer model of **Fig. 21–8**. In the real ear, we see that the gain depends on frequency and that even at the best frequency the gain is only ~22 dB. The difference between the real and theoretical gains has to do with several nonidealities in the middle ear: (1) The middle ear components have their own mechanical properties, such as stiffness and mass, which must be overcome by the sound pressure at the tympanic membrane before sound pressure can be transferred to the inner ear. (2) The tympanic membrane is not like the rigid plate in **Fig. 21–8B**. Instead, it bends and stretches in a highly frequency-dependent manner, thereby decreasing its efficacy as a coupler of sound pressure (Khanna and Tonndorf, 1972). (3) Other mechanically and acoustically important structures in the ear; for example, the compressibility of the air within the middle ear cavities, load the motion of the tympanic membrane and ossicles and use up some of the pressure increase produced by the middle ear transformer. Indeed, the frequency selectivity of the middle ear mechanics and the tympanic membrane response to sound play a large role in determining

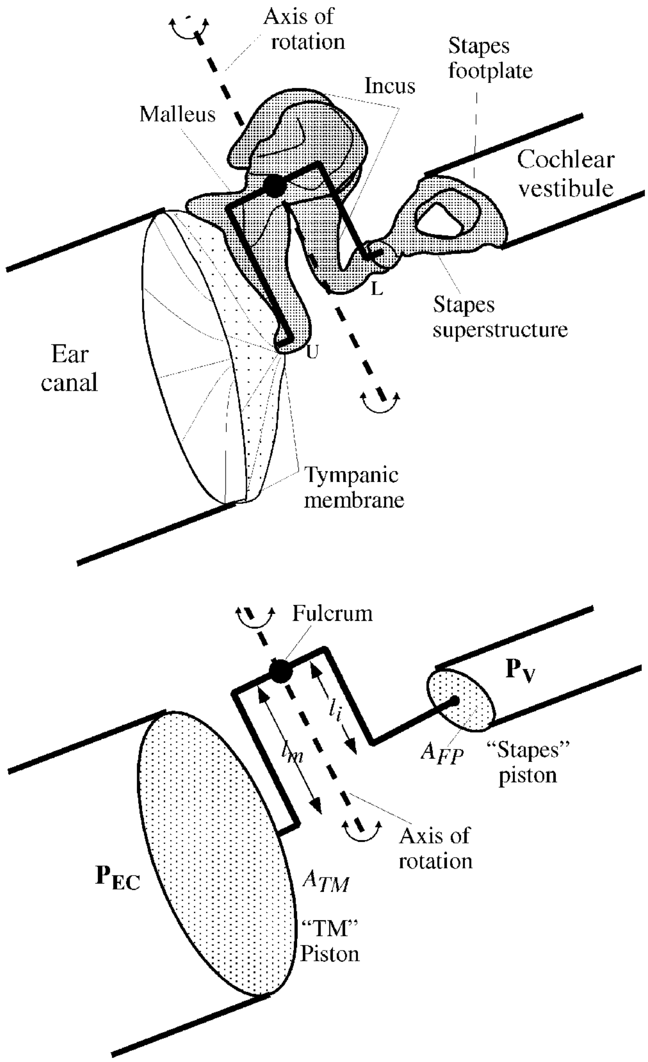


Figure 21–8 ***(A)*** *A schematic of the tympano-ossicular system and* ***(B)*** *a mechanical analog depicting the important structures in the transformation of sound pressure from the external ear (P_{EC}) to the vestibule of the inner ear P_V. A mechanoacoustic transformer increases either pressure or velocity while decreasing the other to maintain the same power output. The differences in the area of the tympanic membrane (A_{TM}) and stapes footplate (A_{FP}) and the different lengths of the ossicular arms (l_m and l_i) act to increase sound pressure at the footplate at the expense of particle velocity.*

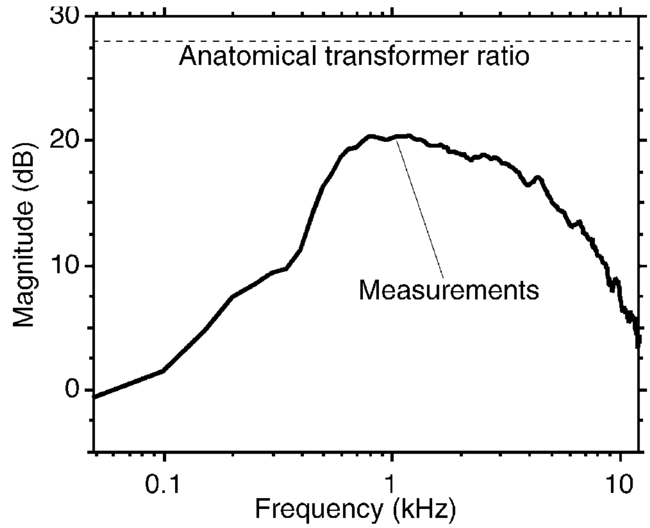


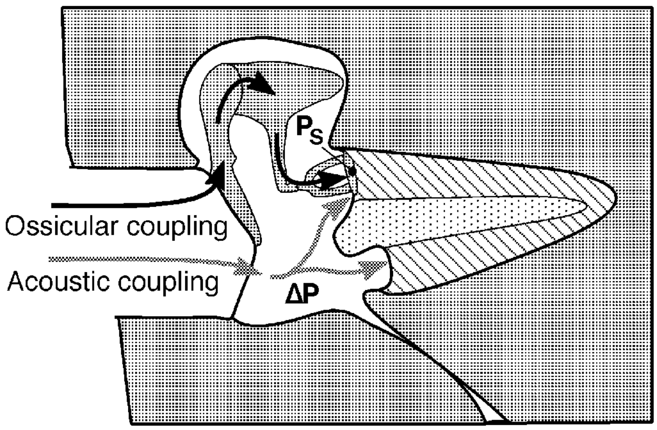
Figure 21–9 *Measurements of the increase in magnitude of sound pressure in the vestibule of the inner ear over the sound pressure at the tympanic membrane (TM) performed in temporal bones (Puria et al, 1997). The curve represents the mean of five measurements. The dashed line at the top shows the theoretical transformer ratio produced by the TM–footplate area ratio and the ossicular lever.*

the efficacy of how sounds of different frequencies stimulate the inner ear.

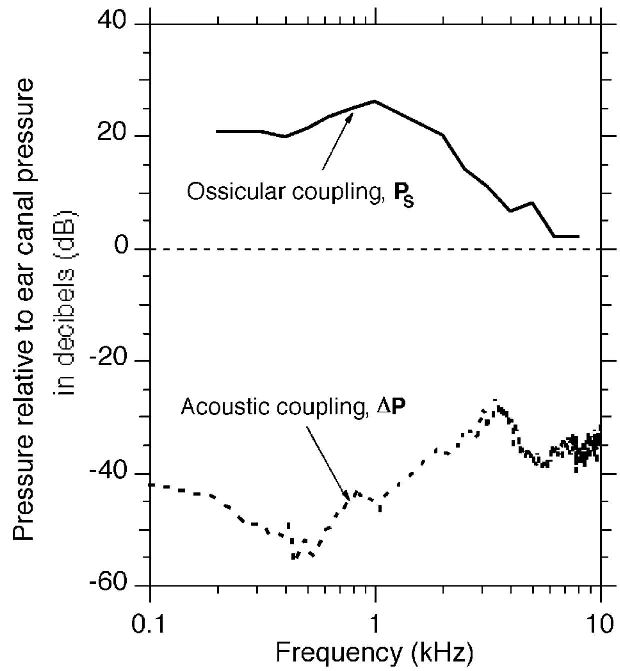
Sound Stimulation of the Inner Ear: Other Paths for Sound Stimulation

The inner ear can be stimulated by sound through paths other than motion of the ossicles, but before we can compare the different sound pathways, we need to understand how sound stimulates the inner ear. The work of von Békésy (1960), Wever and Lawrence (1954), and others (e.g., Voss et al, 1996) suggests that the bulk motion of intracochlear fluid is the stimulus to the inner ear. The inner ear is a fluid-filled space surrounded for the most part by immovable and incompressible bone. The exceptions to these immovable boundaries include the connections to the middle ear (the oval and round windows), and relatively long and narrow fluid-filled connections to the semicircular canals, brain, and body fluid spaces. However, the latter fluid-filled connections are generally thought to play no role in sound passage into the inner ear. Motions of the ossicles drive the stapes footplate, as well as the coupled incompressible cochlear fluid, cochlear structures, and round window, and it is the presence of the round window and its compliant covering membrane that permits the motions of the footplate. When the stapes is pushed in, the round window pushes out. Because of this coupling of the oval and round window by the cochlear fluid, effective stimulation to the inner ear depends on a difference in the sound pressure at the two cochlear windows (Voss et al, 1996; Wever and Lawrence, 1954).

We have already discussed how ossicularly coupled (**Fig. 21–10A**) sounds are transmitted from the ear canal to the oval window; **Fig. 21–10B** shows a different measure of ossicular-conducted sound made by quantifying the effective sound pressure produced on the stapes by the middle ear. By this measure, the ossicular pathway produces as much as 25 dB gain, thereby enhancing both the sound pressure stimulus (P_S) it couples to the oval



A



B

Figure 21–10 *A description and comparison of the ossicular and acoustical sound signals that are coupled to the inner ear by the middle ear.* ***(A)*** *A schematic describing the two paths. Ossicular coupling is produced by the coupled motion of the tympanic membrane, ossicles, and stapes footplate. Acoustical coupling results from the middle ear sound pressure that is produced by the ear canal sound pressure and the motion of the tympanic membrane. (After Merchant et al, 1998.)* ***(B)*** *The middle ear gain data are those of Kurokawa and Goode (1995), and the measurements of the window pressure difference are those of Voss (1998).* P_S*, sound pressure stimulus at the stapes;* Δ*P, difference in sound pressures.*

window and the difference in sound pressures (Δ*P*) between the oval and round window.

The acoustic coupling of ear canal sound to the inner ear results from the direct stimulation of the cochlear windows by sound pressures in the middle ear air spaces. The middle ear sound pressures result from the change in middle ear volume associated with the inward and outward motions of the tympanic membrane. Because the wavelengths of most audible sound frequencies are large compared with the dimensions of the middle ear cavity, the sound within the cavity is fairly homogeneous, and the sound pressure at the oval window approximates the sound pressure at the round window. However, small differences between the magnitudes and phases of the two window pressures result in a small but measurable difference in sound pressure (Δ*P*) at the two windows. In the normal ear, the magnitude of this pressure difference is small compared with ossicularly coupled sounds (**Fig. 21–10B**), and ossicular coupling dominates normal middle ear function (Merchant et al, 1998; Peake et al, 1992).

Another possible path for sound to reach the inner ear is by external sounds vibrating the bones and tissues of the head; however, this pathway is not well described.

THE ACOUSTICS AND MECHANICS OF DISEASED MIDDLE EARS

We can use some of the concepts discussed in the previous section to understand sound transmission in some abnormal states of the middle ear.

Ossicular Interruption with an Intact Tympanic Membrane

When there is ossicular interruption in the presence of an intact drum, ossicular coupling is lost, and sound input to the cochlea results solely from acoustic coupling (Peake et al, 1992). **Fig. 21–10** demonstrates that acoustic coupling is ~60 dB smaller than ossicular coupling; therefore, one would predict that ossicular interruption results in a conductive hearing loss of 60 dB. This prediction is consistent with clinical observations. As shown in **Fig. 21–11A**, there is good agreement between this prediction and the actual (mean) air–bone gap as measured in eight surgically confirmed cases of total interruption of the incudostapedial joint with an intact tympanic membrane (TM). Note that the consistency of the clinical results with the model of acoustic

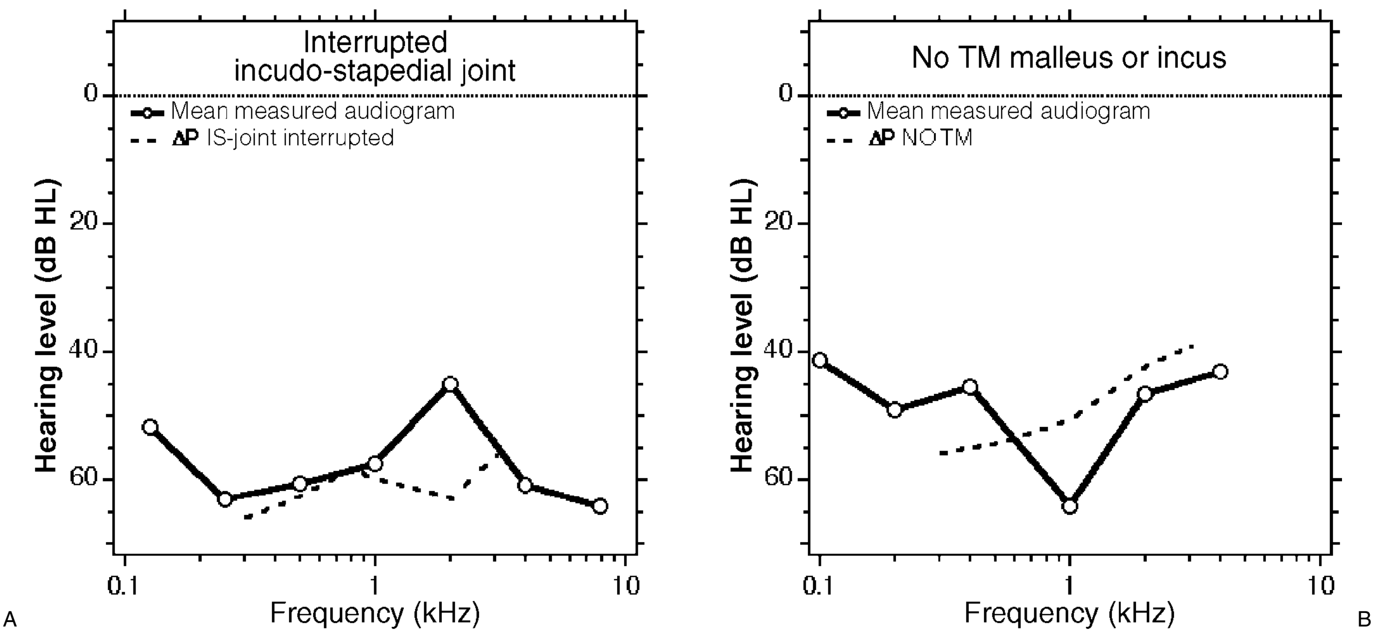


Figure 21–11 *Comparisons of audiometric measurements of hearing function after ossicular disruptions and estimates of hearing function based on measurements of acoustic coupling. (After Peake et al, 1992.)* ***(A)*** *Mean of eight audiograms from patients with surgically confirmed incudostapedial disarticulation, and measurements of ossicular coupling of Békésy (1960).* ***(B)*** *Mean of five audiograms from ears with no tympanic membrane, malleus, or incus (Békésy, 1960) and the acoustic coupling estimates for ears with no tympanic membrane (Peake et al, 1992). IS, incudostapedial; TM, tympanic membrane.*

coupling suggests that stimuli reaching the inner ear through bone conduction mechanisms in this particular condition are small enough to be ignored. This conclusion is supported by other measurements of hearing function in cases where the ossicular route is greatly reduced.

Loss of the Tympanic Membrane, Malleus, and Incus

Loss of the TM, malleus, and incus abolishes ossicular coupling and leads to an enhancement of acoustic coupling of ~10–20 dB as compared with the normal ear (Peake et al, 1992). This enhancement results from the loss of the shielding effect of the TM, which in the normal ear attenuates the middle ear sound pressure by 10 to 20 dB relative to the ear canal sound pressure. Therefore, for the missing TM, malleus, and incus condition, the predicted air–bone gap is 40 to 50 dB. The predicted gap is similar to that measured in patients. **Fig. 21–11B** shows the mean air–bone gap from five cases with total loss of the TM, malleus, and incus). The increase in acoustic coupling due to loss of TM shielding also explains why the hearing of a patient with an interrupted ossicular chain and an intact TM is improved by 10 to 20 dB with the creation of a perforation in the TM.

Ossicular Fixation

Fixation of the stapes footplate due to otosclerosis results in a conductive hearing loss that can range from 5 to 50 dB. Otosclerotic changes in the annular ligament and fixation of the stapes reduce ossicular coupling by hindering stapes motion, resulting in a conductive hearing loss. The amount of loss depends on the degree of decreased stapes motion. For low frequencies, where the stiffness of the annular ligament is a major constraint on the ossicular route in the normal ear, even a small increase in ligament stiffness will produce a measurable conductive loss that is consistent with the initial appearance of a low-frequency hearing loss in otosclerosis.

In contrast, the conductive hearing loss associated with so-called malleus fixation is only 15 to 25 dB. Malleus fixation is generally the result of a bony spur that extends from the epitympanic wall to the malleus head. Because the point of fixation lies near the axis of rotation of the malleus, and because this spur attaches to the malleus head only over a relatively small surface area, it is likely that true "fixation" of the malleus does not occur and some sound is still transmitted across the incudomalleal joint. Hence, so-called malleus fixation leads to a reduction in ossicular coupling, which is consistent with the 15 to 25 dB conductive losses that are observed clinically.

Tympanic Membrane Perforation

Perforations of the TM cause a conductive hearing loss that can range from negligible to 50 dB. The primary mechanism of the air–bone gap due to a perforation is a reduction in ossicular coupling that is caused by a loss in pressure difference across the TM, and the magnitude of the gap is proportional to the degree of reduction in ossicular coupling (Voss, 1998). The perforation-induced loss in ossicular coupling is generally greater at lower frequencies and greater with larger perforations. The perforation also leads to an increase in acoustic coupling by 10 20 dB, because of loss of the shielding effect of the intact TM. The increase in acoustic coupling allows one to predict that the maximum conductive loss from a perforation will be no more than 40 to 50 dB, which is consistent with clinical observations.

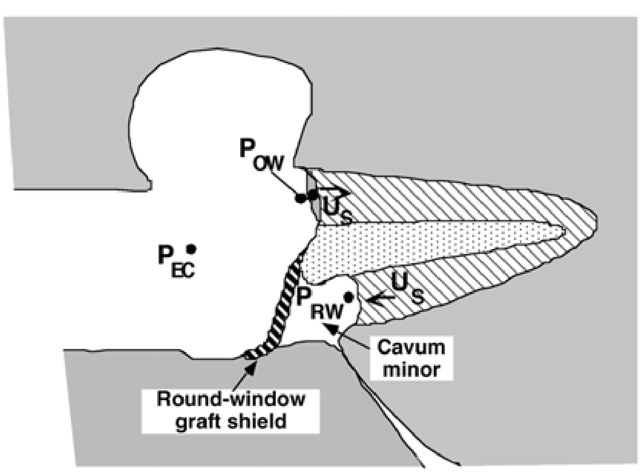
Middle Ear Effusion

Fluid in the middle ear, a primary feature of otitis media with effusion (OME), is associated with a conductive hearing loss of up to 30 to 35 dB, though the degree and frequency dependence of individual losses can vary. The conductive loss occurs because of reduction in ossicular coupling due to several mechanisms. At high frequencies >1 kHz, hearing loss is caused primarily by mass loading of the TM by fluid, with decreases in sound transmission of up to 20 to 30 dB. The effect increases as more of the TM is covered with fluid. At low frequencies <1 kHz, hearing loss is due to an increase in impedance of the middle ear air space (resulting from reduced middle ear air volume) and negative middle ear static pressure (which is often associated with OME). It appears that viscosity of the fluid does not significantly influence middle ear sound transmission in OME.

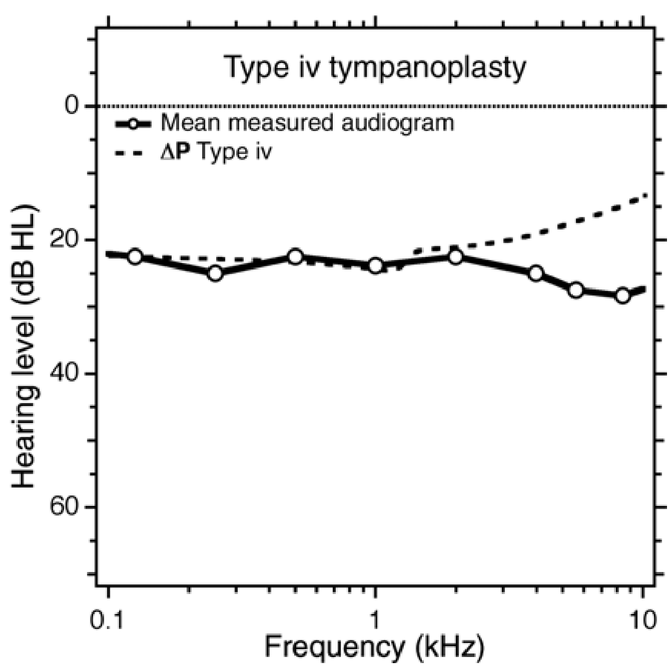
THE ACOUSTICS AND MECHANICS OF RECONSTRUCTED MIDDLE EARS

Tympanoplasty Techniques without Ossicular Linkage: Types IV and V

When the TM and ossicles are missing and there is a canal wall-down mastoid cavity, a type IV tympanoplasty is a surgical option. Incoming sound from the ear canal is allowed to impinge directly onto the stapes footplate, while the round window is shielded by a tissue graft such as temporalis fascia (**Fig. 21–12A**). If the stapes footplate is ankylosed, it is removed and replaced by a fat graft, and this arrangement constitutes a type V tympanoplasty. In both type IV and type V procedures, there is no ossicular coupling, and residual hearing depends solely on acoustic coupling (Peake et al, 1992. The introduction of a tissue graft to acoustically



A



B

Figure 21–12 ***(A)*** *A schematic of a type IV tympanoplasty. P_{EC}, P_{OW}, and P_{RW} denote the sound pressures in the ear canal, outside the oval window and outside the round window. U_S is the volume velocity of the stapes.* ***(B)*** *A comparison of the mean of the five best hearing results of Wullstein (1960) after type IV tympanoplasty and predictions based on the loss of ossicular coupling and the enhancement in acoustic coupling produced by the graft shield. (After Merchant et al, 1998.)*

shield the round window enhances acoustic coupling. Under optimal conditions, one can achieve "maximum" acoustic coupling, and one would predict an air–bone gap of only 20 dB. This prediction is consistent with the best type IV hearing results observed clinically (**Fig. 21–12B**). The optimal conditions required to maximize acoustic coupling include normal footplate mobility, a sufficiently stiff acoustic shield, and aeration of the round window niche (Merchant et al, 1995; Peake et al, 1992). In clinical practice, the surgeon should (1) preserve normal stapes mobility by preferably covering the footplate with a very thin split thickness skin graft, not a fascia graft; (2) reinforce a fascia shield with cartilage or Silastic, 1 mm thick; (3) create conditions that promote aeration of the round window niche; and (4) preserve mobility of the round window membrane.

The interposition of a graft to acoustically shield the round window in a type IV or type V tympanoplasty primarily reduces the magnitude of the sound pressure at the round window. Under these circumstances, acoustic coupling (i.e., the difference in sound pressures at the oval and round window) is relatively insensitive to the phase difference between the two pressures (Merchant et al, 1998). In other words, the difference in magnitude between the oval and round window pressures is more important than the difference in phase in determining the effective cochlear input.

Tympanoplasty Techniques with Preservation of Ossicular Linkage: Types I, II, and III

Types I, II, and III tympanoplasty involve reconstruction of the TM and/or the ossicular chain. The final hearing results depend on three factors: the efficacy of the reconstructed TM, the efficacy of the reconstructed ossicular chain, and the adequacy of middle ear aeration and its static pressure. At the present time, our knowledge of the mechanics of these three factors is incomplete, and much work remains to be done.

Tympanic Membrane Reconstruction

The TM is the major transducer of the middle ear. Motion of the normal TM is complex (Khanna and Tonndorf, 1972). At frequencies below 1000 Hz, the entire TM moves in-phase, but the magnitude of motion varies with location. At higher frequencies, different parts of the TM move with varying magnitudes and phases. Clinical observations suggest that those surgical techniques that tend to restore or preserve the normal anatomy of the drum can lead to good hearing results. However, the acoustical and mechanical properties of reconstructed TMs are not well understood at the present time, and more research is needed. For example, (1) little is known of the mechanical significance of the arrangements of structural fibers in the TM; (2) although it has been argued that the conical shape of the normal tympanic membrane plays a role in middle ear function, the possible effects of changes in TM shape on middle ear function are not understood; and (3) although many existing models of TM function have been shown to fit some of the available data (e.g., Funnell and Decraemer, 1996), there are wide differences in the structure of these models, and little effort has been made to compare their significant similarities and differences.

Ossicular Reconstruction

Ossiculoplasty is required in 40 to 90% of all tympanoplasties. A wide variety of ossicular grafts and prostheses are in use. Currently, there are few scientific data on the precise structure–function relationship concerning their acoustical properties. Factors that can influence the acoustic performance of an ossicular prosthesis include its stiffness, mass, position, tension, and coupling (Merchant et al, 1998). In general, the stiffness of a prosthesis will not be a significant factor as long as the stiffness is much greater than that of the stapes-cochlear impedance. For clinical purposes, prostheses made of ossicles, cortical bone, and most synthetic materials generally meet this requirement.

Analysis of a middle ear model suggests that an increase in ossicular mass should not cause significant detriment in sound transmission. Increases in the mass of a prosthesis up to 16 times that of the stapes mass (which is 3 mg) were predicted to produce <10 dB air–bone gaps and only at frequencies above 1 kHz. Using a temporal bone preparation, it has been reported that a mass of 5 mg added to the stapes head led to a 13 to 15 dB loss at high frequencies (Goode and Nishihara, 1994). Thus the acoustic effects of mass of an ossicular strut need to be clarified. Measurements in a human temporal bone preparation have suggested that the angle between the stapes and a prosthesis should be <45 degrees for optimal sound transmission. Although it is ideal to attach a prosthesis to the manubrium, temporal bone data (Goode and Nishihara, 1994) have shown that acceptable results can occur with a prosthesis placed against the posterosuperior quadrant of the TM as long as 3 to 4 mm of the diameter of the prosthesis contacts the TM.

Tension and length are critical in determining the hearing result. The mechanical impedance of biological structures is inherently nonlinear, and measurements (e.g., tympanometry) have shown that the TM and the annular ligament act as linear elements only over the range of small motions ($<10\ \mu$m) associated with physiological sound levels. The larger displacements produced by a prosthesis that is too long will produce a stiffening of the annular ligament and TM, resulting in excessive tension, and hence, a poor hearing result.

Coupling refers to how well a prosthesis "adheres" to the footplate or TM, and the degree of coupling will determine whether or not there is slippage in sound transmission at the end of a prosthesis. Thus a prosthesis will transmit sound effectively only if there is "good" coupling at both ends. Clinical observations indicate that it is rare to obtain a firm union between a total ossicular replacement prosthesis (TORP) and the stapes footplate, and hence, inadequate coupling at the TORP–footplate joint may be an important cause of a persistent postoperative air–bone gap. The physical factors that control coupling have not been determined in a quantitative manner, and further study of this parameter is warranted.

Middle Ear Aeration and Static Pressure

An important parameter for the success of tympanoplasty is the presence of an aerated middle ear. Aeration allows the TM graft, the ossicles, and the round window to move. Nonaerated ears have large 45 to 60 dB air–bone gaps, whereas in an adequately aerated ear, the size of the air–bone gap is directly proportional to the efficiency of the TM and ossicles. How much air is necessary behind the TM (i.e., within the middle ear and mastoid)? Model analyses of the effects of varying the volume of the middle ear and mastoid predict an increasing low-frequency hearing loss as volume is reduced. The normal, average volume of the middle ear and mastoid is 6 cc; a combined middle ear and mastoid volume of 0.5 cc was predicted to result in a 10 dB conductive loss. Volumes smaller than 0.5 cc were predicted to lead to progressively larger gaps, whereas volumes greater than 1.0 cc were predicted to provide little additional acoustic benefit. Experimental studies in our laboratory using a human temporal bone preparation where the middle ear and mastoid volume was reduced progressively have shown results consistent with the model prediction. Another parameter that can influence middle ear sound transmission is the static air pressure in the middle ear space. Experiments in a human temporal bone preparation have demonstrated that middle ear static pressure can have different effects on sound transmission at different frequencies (Murakami et al, 1997). The mechanisms and structural alterations by which pressure changes reduce middle ear sound transmission are not well defined, and possible sites of pressure sensitivity include the TM, the annular ligament, the incudomalleal joint, and the suspensory ligaments of the ossicles. Some of these physical structures are drastically altered as a result of tympanoplasty, and the acoustic effects of negative and positive middle ear static pressure in such ears need to be characterized.

Canal Wall-Up versus Canal Wall-Down Mastoidectomy

A canal wall-down mastoidectomy poses several considerations from an acoustical and mechanical perspective (when compared with a canal wall-up procedure): First, the canal wall-down procedure results in a significant reduction in the size of the residual middle ear air space. However, as long as this air space is greater than 0.4 cc, the acoustical effect of loss of sound transmission should be less than 10 dB. Second, a canal wall-down procedure results in the creation of a large air space lateral to the TM (i.e., the air space within the mastoid bowl including the external auditory canal). This mastoid bowl and ear canal air space generates resonances that can influence middle ear sound transmission favorably or unfavorably. For example, Goode et al (1977) described a patient with a 4.1 cc radical mastoid cavity that had a 25 to 30 dB resonance for frequencies between 2500 and 4000 Hz, in whom these resonances were felt to contribute to the patient's good hearing at these higher frequencies. However, the structure–function relationships between the size and shape of the normal and modified mastoid cavity and cavity resonances have not been well defined. An improved understanding of this issue should help to develop guidelines for an otologic surgeon to configure mastoid cavities in ways that could be acoustically beneficial. Third, after a canal wall-down procedure, the TM graft comes to lie in a more medial position compared with normal, and the TM graft is made to couple to the stapes head or to a prosthesis such as a TORP. The mechanics of such a TM graft and its coupling to the stapes/TORP are likely to be different from normal and need to be characterized. For example, when placed against the stapes head, such a graft is often in contact with the promontory and tympanic segment of the facial canal; such contact may limit the functional surface area of the graft. When a TORP is considered, if the top of the TORP rises much above the level of the oval window niche, the TORP tends to extrude; if the TORP is kept low, the prosthesis often settles against the margins of the niche.

Stapedotomy

The output of the middle ear can be quantified by the volume velocity of the stapes, where this volume velocity is equal to the stapes linear velocity multiplied by the area of the stapes footplate. After a stapedotomy, the effective area of the footplate is reduced to the area of the prosthesis, thereby reducing the volume velocity produced by a given stapes (linear) velocity. The reduction in effective footplate area also reduces the area of the cochlear fluid over which the force generated by the stapes is applied. Whereas the reduced footplate area leads to a local increase in pressure over the surface of the prosthesis area, the average pressure at the cochlear entrance is reduced. The reduction in stapes volume velocity and cochlear sound pressure leads to a decrease in ossicular coupling and the development of an air–bone gap. The smaller the area of the stapes prosthesis, the greater the air–bone gap. Model predictions of the relationship between piston diameter and residual air–bone gap after a stapedotomy suggest that a 0.8 mm piston causes a 5 dB gap at frequencies of 1000 Hz and below, a 0.6 mm piston is predicted to cause a 10 dB gap, and 0.4 mm piston is predicted to cause a 15 dB gap. These predictions assume that the effective vibrating "footplate" surface area after a stapedotomy is no more than the area of the lower end of the prosthesis. In cases of partial stapedectomy with placement of a vein graft and a stapes prosthesis, the effective vibrating surface may be greater than the area of the prosthesis alone, and our predictions may overestimate the air–bone gap.

SUMMARY

This chapter has summarized some simple acoustic ideas, such as how sound propagates, what sound pressure is, and how the propagation of sound around objects depends on the relationship between the dimensions of the object and the wavelength of the sound. The influence of sound propagation velocity and wavelength on how we localize sounds is described. Also discussed is how the mechanics and acoustics of the middle ear shape the sounds that reach the inner ear and how pathological alterations in middle ear structure can affect hearing. Finally, the acoustics and mechanics after middle ear reconstructive surgery are discussed.

Suggested Readings

American National Standards Institute. Audiometer Standard 1969;3:6

Dallos P. The Auditory Periphery. New York: Academic Press; 1973

Funnell WRJ, Decraemer WM. On the incorporation of moiré shape measurements in finite-element models of the cat eardrum. J Acoust Soc Am 1996;100:925–932

Goode RL, Friedrichs R, Falk S. Effect on hearing thresholds of surgical modification of the external ear. Ann Otol Rhinol Laryngol 1977;86:441–451

Goode RL, Nishihara S. Experimental models of ossiculoplasty. Otolaryngol Clin North Am 1994;27:663–675

Khanna SM, Tonndorf J. Tympanic membrane vibrations in cats studied by time-average holography. J Acoust Soc Am 1972; 51:1904–1920

Kurokawa H, Goode RL. Sound pressure gain produced by the human middle ear. Am J Otol 1995;113:349–355

Merchant SN, Ravicz ME, Voss SE, Peake WT, Rosowski JJ. Toynbee Memorial Lecture 1997: middle ear mechanics in normal, diseased and reconstructed ears. J Laryngol Otol 1998;112:715–731

Merchant SN, Rosowski JJ, Ravicz ME. Middle-ear mechanics of type IV and type V tympanoplasty, II: Clinical analysis and surgical implications. Am J Otol 1995;16:565–575

Murakami S, Gyo K, Goode RL. Effect of middle ear pressure change on middle ear mechanics. Acta Otolaryngol (Stockh) 1997;117:390–395

Peake WT, Rosowski JJ, Lynch TJ III. Middle-ear transmission: acoustic vs. ossicular coupling in cat and human. Hear Res 1992;57:245–268

Puria S, Peake WT, Rosowski JJ. Sound-pressure measurements in the cochlear vestibule of human cadavers. J Acoust Soc Am 1997;101:2745–2770

Rosowski JJ. The effects of external- and middle-ear filtering on auditory threshold and noise-induced hearing loss. J Acoust Soc Am 1991;90:124–135

Rosowski JJ. Models of external and middle-ear function. In: Auditory Computation New York: Springer-Verlag; 1996:15–61. Hawkins H, McMullen T, Popper A, Fay R, eds. The Springer Handbook of Auditory Research; vol 6

Shaw EAG. The external ear. In: Auditory System (pp. 455–490). New York: Springer-Verlag; 1974:455–490. Keidel WD, Neff WD, eds. Handbook of Sensory Physiology; vol 5

Sivian LJ, White SD. On minimum sound audible fields. J Acoust Soc Am 1933;4:288–321

von Békésy G. Experiments in Hearing. New York: McGraw-Hill; 1960

Voss SE. Effects of Tympanic-Membrane Perforations on Middle-Ear Sound Transmission: Measurements, Mechanisms and Models [PhD dissertation]. Cambridge: Massachusetts Institute of Technology; 1998

Voss SE, Rosowski JJ, Peake WT. Is the pressure difference between the oval and round windows the effective acoustic stimulus for the cochlea? J Acoust Soc Am 1996;100(3):1602–1616

Wever EG, Lawrence M. Physiological Acoustics. Princeton, NJ: Princeton University Press; 1954

Wullstein H. Results of tympanoplasty. Arch Otolaryngol 1960; 71:478–485

Yost WA. Fundamentals of Hearing. New York: Academic Press; 1994

Zwislocki J. The role of the external and middle ear in sound transmission. In: Human Communication and Its Disorders. New York: Raven Press; 1975:45–55. Tower DB, ed. The Nervous System; vol 3

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Which of the following statements best describes the middle ear gain in humans?
   - A. 30 dB and independent of frequency
   - B. 20 dB and independent of frequency
   - C. 20 dB with a best frequency near 4 to 8 kHz
   - D. 20 dB with a best frequency near 1 to 2 kHz
2. The inner ear responds best to
   - A. Sound pressure applied to the oval window
   - B. Sound pressure applied to the round window
   - C. The sum of sound pressures at the two windows
   - D. The difference in sound pressure at the two windows
3. What is the primary mechanism of conductive hearing loss due to a tympanic membrane perforation?
   - A. A reduction in area of the tympanic membrane
   - B. A reduction in coupling of the tympanic membrane motion to the malleus
   - C. A reduction in sound pressure difference across the tympanic membrane
   - D. An increase in sound pressure at the round window compared with the oval window
4. Which of the following statements concerning hearing results after a stapedotomy are false?
   - A. The smaller the area of the stapes prosthesis, the smaller the air–bone gap.
   - B. The larger the area of the stapes prosthesis, the smaller the air–bone gap.
   - C. The area of the stapes prosthesis is not correlated to the size of the air–bone gap.
   - D. Model analyses predict a small residual air–bone gap after a stapedotomy.

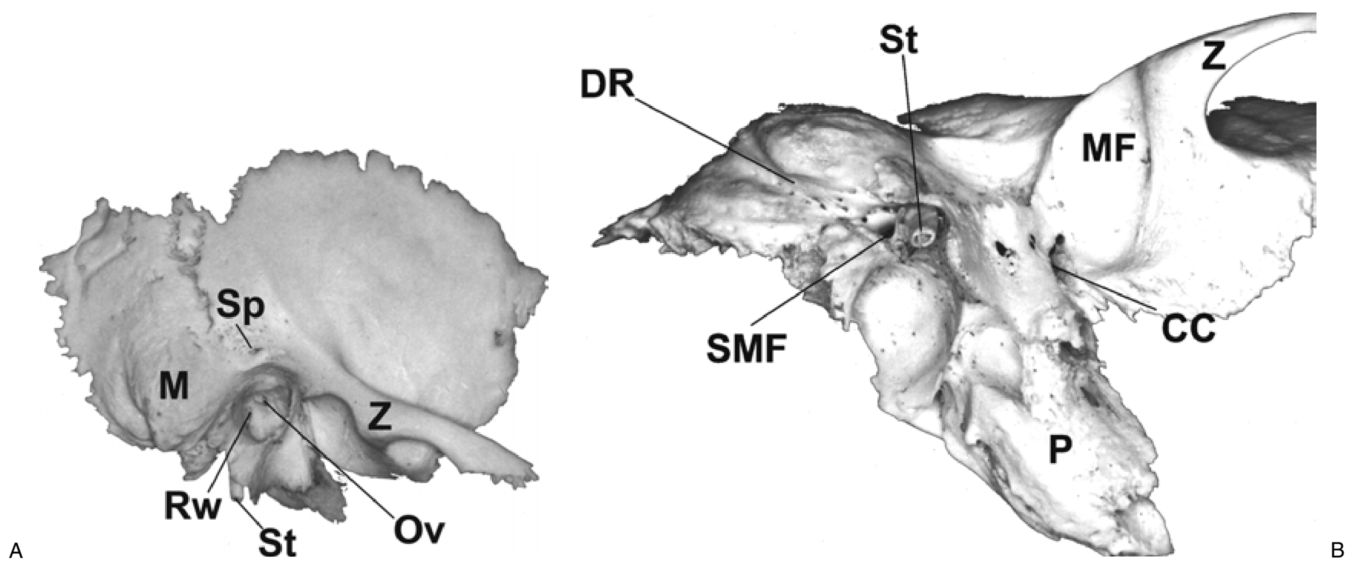
Chapter 22
SURGICAL ANATOMY OF THE TEMPORAL BONE

HINRICH STAECKER AND ADRIEN A. ESHRAGHI

The anatomy of the temporal bone represents the junction of the structures of the neck, nasopharynx, and the cranium. Multiple vital neural and vascular structures pass through this area. In addition the temporal bone contains the sensory receptor organs of hearing and balance. Inferiorly, the sternocleidomastoid muscle and the great vessels of the neck attach and enter the temporal bone. Anteriorly, the temporal bone articulates with the condyle of the mandible and attaches to the midface via the zygomatic arch. The eustachian tube connects the aerated spaces of the middle ear and mastoid cavity to the nasopharynx. Medially, the temporal bone contacts the posterior and middle fossae of the skull base. Being at such a nexus makes the temporal bone a target for a bewildering variety of diseases. Although the anatomy of the temporal bone will be described in this passage, a full understanding of the complex anatomy of the temporal bone can only be obtained through repeated anatomical dissection.

EXTERIOR ANATOMY OF THE TEMPORAL BONE

The lateral view of the temporal bone (**Fig. 22–1A**) presents several key anatomical features. The temporal line represents the insertion point of the temporalis muscle and roughly approximates the position of the floor of the middle cranial fossa. The mastoid tip is the inferior extent of the mastoid portion of the temporal bone. Its inferior extent varies with age, with infants having only a minimal projection of the mastoid. This, along with the delayed ossification of the tympanic ring, makes it necessary to modify incisions for mastoid surgery in infants to avoid damaging the facial nerve. The spine of Henle approximates the position of the antrum. Visible in the external auditory canal are the tympanosquamous and tympanomastoid fissures that mark the fusion planes between the squamous, tympanic, and mastoid bones. The petrous bone makes up the fourth component of the temporal bone and lies deep to the lateral portion of the temporal bone. From the inferior aspect (**Fig. 22–1B**) one can observe the relationship between the stylomastoid foramen and the digastric muscle that inserts in a groove on the mastoid tip. Internally, this represents the digastric ridge. The carotid canal, mandibular fossa, and styloid process are also present. The superior projection encompasses the middle cranial fossa. This area is rather poor in landmarks. The arcuate eminence approximates the position of the superior semicircular canal. On the posterior projection of the temporal bone (**Fig. 22–1C**)



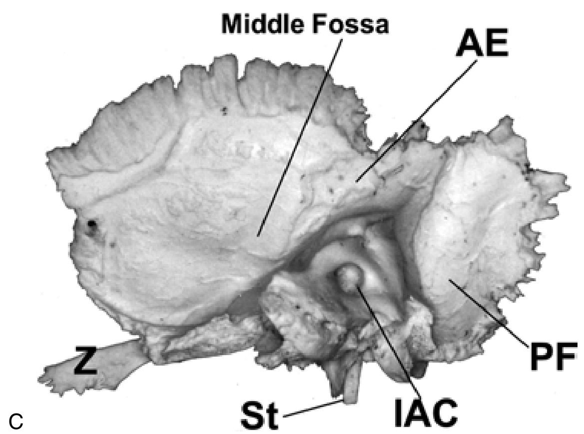


Figure 22–1 *External bony structures of the temporal bone:* ***(A)*** *lateral view,* ***(B)*** *inferior view, and* ***(C)*** *posteromedial view. AE, arcuate eminence; CC, carotid canal; DR, digestive ridge; IAC, internal auditory canal; M, mastoid; MF, mandibular fossa; Ov, oval window; P, petrosquamous suture; PF, petrotympanic fissure; Rw, round window niche; SMF, stylomastoid foramen; Sp, spine of Henle; St, styloid process; Z, zygoma.*

one can see the internal auditory canal (IAC). The cochlear nerve occupies the anteroinferior portion of the IAC, and the facial nerve occupies the anterosuperior quadrant of the IAC and is separated from the cochlear nerve via the transverse crest (see **Fig. 22–2**). The superior and inferior vestibular nerves occupy the posterior half of the IAC. Of note is that the IAC is in line with the external auditory canal (EAC), providing a useful radiological and surgical landmark. The operculum is the opening to the bony portion of the endolymphatic sac. The transverse and sigmoid sinuses can be seen coursing along the edge of the temporal bone.

ANATOMY OF THE TYMPANIC MEMBRANE AND MIDDLE EAR CAVITY

The tympanic membrane is an irregularly round membrane that measures ~8 × 9 mm and is ~0.1 mm thick. It is divided into a pars tensa and a pars flaccida that are separated by the attachment of the anterior and posterior tympanic striae. The pars tensa is firmly attached to the annulus and in its center is attached to the long process of the malleus. The pars flaccida forms the lateral border of Prussak's space. The pars flaccida at this point attaches to the tympanic incisura or notch of Rivinus, thereby making up the superior recess of the tympanic membrane. Prussak's space is defined superiorly by the lateral malleal ligament, medially by the head of the malleus, and inferiorly by the anterior and posterior malleal folds. The pars tensa consists of three layers: a squamous epithelial external layer, a fibrous middle layer (arranged in both radial and circular patterns), and a mucosal internal layer. The pars flaccida is composed of two layers, and the fibrous middle layer is not present.

The middle ear cavity (MEC) itself can be divided into five parts: the hypotympanum, compromising the floor of the middle ear; the epitympanum, making up the roof; the protympanum and posterior tympanum, composing the anterior and posterior borders, respectively; and the mesotympanum, making up the central

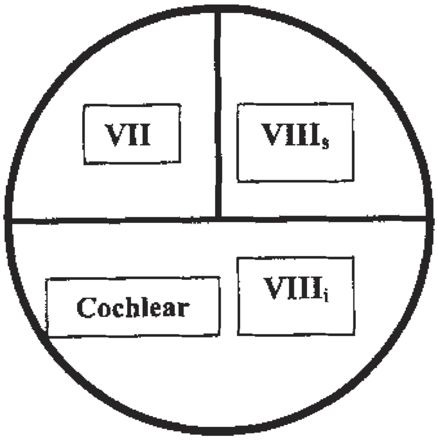


Figure 22–2 *The organization of the cranial nerves within the internal auditory canal (IAC) (medial to lateral view). The seventh nerve is located in the anterosuperior segment of the IAC and is separated from the superior division of the vestibular nerve (VIII$_s$) in the distal portion of the IAC by Bill's bar. The transverse crest separates the distal IAC into an upper and lower portion, providing another important surgical landmark. The cochlear nerve occupies the anteroinferior portion of the canal, whereas the inferior division of the vestibular nerve (supplying the macula of the saccule and the posterior canal crista) occupies the inferior portion of the IAC.*

section (see **Fig. 22–3**, canal wall-up mastoidectomy view of the MEC). Sound conduction occurs through the action of three ossicles, the malleus, incus, and stapes (see **Fig. 22–4**, the isolated ossicles that form the ossicular chain).

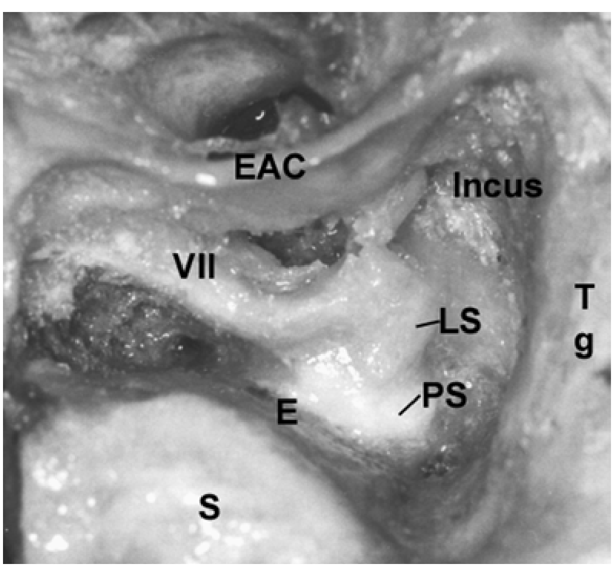


Figure 22–3 *Temporal bone dissection showing a canal wall-up mastoidectomy. E, endolymphatic sac; EAC, external auditory canal; LS, lateral semicircular canal; PS, posterior semicircular canal; S, sigmoid sinus; Tg, tegmen; VII, facial nerve.*

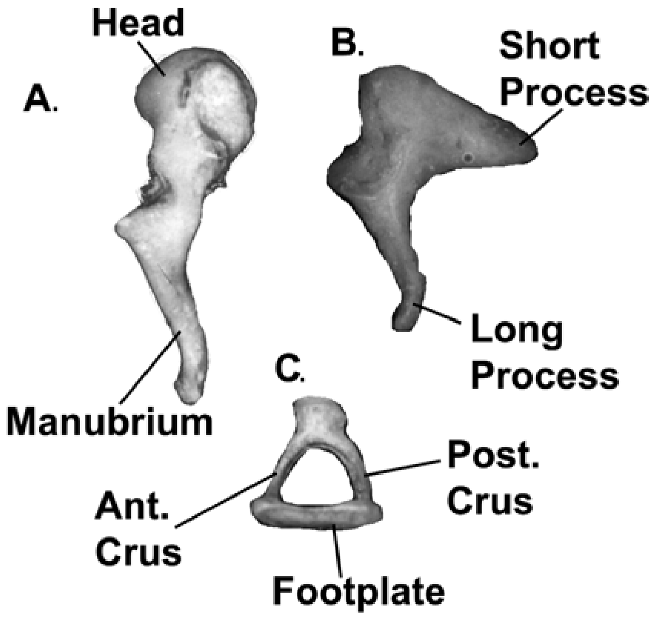


Figure 22–4 *The ossicles of the middle ear cavity:* ***(A)*** *malleus,* ***(B)*** *incus, and* ***(C)*** *stapes.*

The malleus is divided into the head, neck, lateral process, anterior process, and manubrium (**Fig. 22–4A**). The anterior process runs toward the glasserian fissure (pertotympanic fissure), the exit point of the chorda tympani nerve from the middle ear. The malleus is attached to the tympanic membrane via a cartilaginous cap at the lateral process and is completely enveloped in the lamina propria layer at the level of the umbo. Removal of the tympanic membrane at this point is impossible without causing a perforation. Just inferior to the lateral process, the malleus is loosely attached to the tympanic membrane via the plica mallearis. This layer may be elevated to allow for the attachment of an ossicular prosthesis to the malleus.

The incus is composed of the body, which articulates with the head of the malleus, the short process, and the long process, which articulates with the head (caput) of the stapes (**Fig. 22–4B**).

The stapes is the final bone in the ossicular chain that links the tympanic membrane to the oval window and that on average measures 3.2 mm in height (**Fig. 22–4C**). The footplate generally measures an average size of 1.4 × 2.9 mm and is articulated into the oval window via the annular ligament.

The ossicles are suspended by a complex series of membranes and ligaments that are thought to be formed by the migration and interaction of the mesenchymal tissues of the four embryonic pouches. These ligaments and folds of the mesotympanum and epitympanum constitute barriers that tend to channel disease growth (e.g., cholesteatoma) along characteristic anatomical pathways. Besides the anterior and posterior malleal folds, the malleus is suspended by the anterior, lateral, and superior suspensory ligaments. The tensor tympani tendon attaches the neck of the malleus medial to the cochleariform process and to the tensor tympani muscle. The incus is held in place by the posterior incudal ligament that attaches to the short process of the incus and by the medial and lateral incudal ligaments that secure the body of the incus to the head of the malleus. The long process of the incus articulates with the head of the stapes and is the most common site for bony erosion in chronic otitis media. The tendon of the stapedius attaches to either the neck or the posterior crus of the stapes. The stapedius muscle is the smallest muscle of the body and is innervated by a branch of the facial nerve. The body of the stapedius muscle is encased within the pyramidal eminence of the middle ear. Lateral to the pyramidal eminence, the chorda tympani nerve enters into the MEC. The MEC is connected to the pneumatized portion of the mastoid bone via the posterior epitympanum that leads into the aditus ad antrum and from there into the mastoid antrum. The carotid artery and the opening to the eustachian tube make up the anterior wall of the middle ear, whereas the roof of the MEC (tegmen tympani) separates the middle ear from the middle fossa and temporal lobe dura. As discussed previously, the lateral boundaries of the MEC are the tympanic ring and membrane, and the scutum (consisting of bone at the junction of the squamous and tympanic bones). Erosion of the scutum is a common radiological sign of cholesteatoma.

The medial wall of the MEC borders the sensory organs of the inner ear. The oval and round windows are separated by the sinus tympani, a niche where cholesteatoma can be overlooked. The subiculum is a ridge of bone that separates the round window niche from the sinus tympani. The ponticulus separates the oval window from the sinus tympani. The facial canal is located superior to the oval window niche and is a common location for dehiscence of the facial nerve. The facial nerve turns medially to form the geniculate ganglion just posteromedially from the cochleariform process. This in turn represents the second genu. The area anterior to the head of the malleus is the epitymanic recess. This may be partially separated by a septum of bone.

One of the most important structures for maintaining adequate function of the middle ear is the eustachian tube. The eustachian tube is composed of a bony portion (one third total length, proximal to the middle ear) and a cartilaginous portion (two thirds total length, distal to the middle ear) ~3 cm long in its total length and ~2 mm wide at its narrowest point. It crosses the carotid artery and opens into the nasopharynx. The cartilaginous portion of the tube is shaped like an upside down U and opens into the fossa of Rosenmüller. Tumors in this region (nasopharynx) may present as a unilateral serous otitis media. The eustachian tube is lined by ciliated pseudostratified respiratory type epithelium. Opening of the tube is an active event controlled by three muscles: the tensor vela palatini, the levator vela palatine, and the salpingopharyngeal. Closure of the tube is a passive event. The dominant muscle is the tensor vela palatine, innervated by cranial nerve (CN) V, that acts to pull the lateral portion of the cartilage inferiorly. The levator veli palatine, innervated by CN X, opens the lumen of the eustachian tube.

ANATOMY OF THE MASTOID AND PETROUS APEX

The mastoid cavity is variable in size, depending on both the age of the patient and any history of prior otologic disease. During infancy the mastoid consists of a small cavity (the antrum). With increasing age, pneumatization of the mastoid air cells develops, yielding a variable-sized cavity. Long-standing eustachian tube dysfunction can prevent pneumatization and results in a sclerotic cavity, which is seen in up to 20% of the population. Several pneumatization tracts connect the mastoid air cells to the hypotympanum (retrofacial cells) and to the petrous apex (posteromedial, posterosuperior, and subarcuate cell tracts). The petrous apex air cells are in turn connected to the middle ear and/or the bony portion of the eustachian tube. The mastoid cavity itself is divided into a lateral and a medial half by the petrosquamous (Körner's) septum. There are also several accessory air cell tracts, including the retrosigmoid cells, the tip cells, the tegmental cells, and the sinodural cells. Awareness of these air cells is important in removing residual chronic otitis media.

ANATOMY OF THE COCHLEA AND VESTIBULAR SYSTEM

The membranous sensory structures of the inner ear are surrounded by a dense bony capsule that is derived from periotic cephalic mesenchyme (**Fig. 22–3**; see Chapter 20). The color and hardness of the bony otic capsule make it easily recognizable during temporal bone surgery. The inner ear may be divided into the cochlea, the vestibule, the three semicircular canals with their associated ducts, and the endolymphatic duct and sac. The oval window opens into the vestibule that contains two bony depressions. The spherical recess directly below the stapedial footplate holds the saccule, and the elliptical recess holds the utricle. The maculae of the saccule and utricle are composed of a field of supporting cells and hair cells that are covered by a gelatinous otolithic membrane that contains many small calcium phosphate crystals (otoconia). These maculae with their otoconial membranes function as both gravity and acceleration receptors, and these two sensory maculae lie at right angles to each other. Posteriorly, the vestibule opens into the semicircular canals. Each semicircular canal contains a semicircular duct that has an ampullated and a nonampullated end as it joins into the utricle. The posterior and superior ducts join together to form a common nonampullated duct called the crus commune. The cristae located within the ampullae are complex sensory receptors that contain secretory cells, supporting cells, and hair cells. The stereociliary bundles of the hair cells are embedded into a gelatinous material called a cupula that is attached to the roof of each ampulla. Angular acceleration results in deflection of the stereocilia, causing a change in the basal firing rate of the vestibular nerve. The superior canal runs through the floor of the middle fossa making up the arcuate eminence, an important landmark for orientation of the surgeon during middle fossa surgery. The cochlea lies anterior to the vestibule and consists of 2½ turns, with a total average length of 32 mm. Within these bony confines the membranous labyrinth (inner ear) consists of a sac containing a high-potassium, low-sodium fluid (endolymph) and is suspended in and surrounded by a fluid-filled space encased in bone that is low in potassium and high in sodium (perilymph). Within the cochlea there are three fluid-filled spaces called the scala media (filled with endolymph) and the scala tympani and the scala vestibuli (both filled with perilymph). The internal fluid spaces of the saccule, utricle, and semicircular ducts contain endolymph. It is Reissner's (vestibular) membrane that separates the scala vestibuli from the scala media and the basilar membrane that separates the scala media from the scala tympani (see **Fig. 22–5**). The round window membrane opens into the basal turn of the scala tympani.

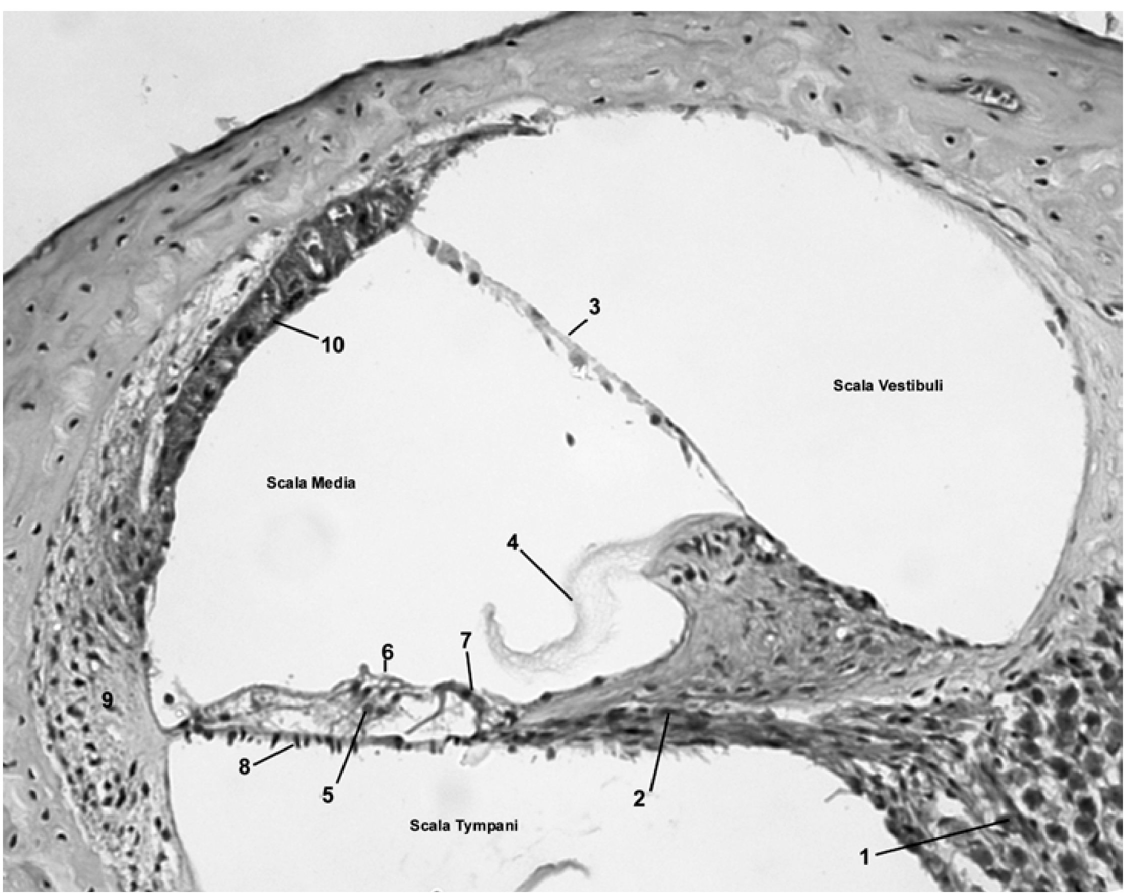


Figure 22–5 *A cross-section view through the cochlear duct of an adult mouse. 1, spiral ganglion; 2, peripheral extensions of the cochlear nerve fibers; 3, Reissner's membrane; 4, tectorial membrane; 5, Deiters' cells; 6, outer hair cells; 7, inner hair cell; 8, basilar membrane; 9, spiral ligament; 10, stria vascularis.*

The communication between perilymphatic spaces of the scala vestibuli and scala tympani occurs at the apex of the cochlea; this area of communication is designated the helicotrema of the cochlea. The lateral cochlear wall consists of the stria vascularis and the spiral ligament, which are both metabolically active structures, responsible for maintaining the ionic balance within the cochlea. The organ of Corti is a complex structure consisting of supporting cells, three rows of outer hair cells, and a single row of inner hair cells. The inner hair cells are the actual receptor cells of hearing, whereas the outer hair cells serve as modulators of hearing and fine-tune the area of the basilar membrane that responds to the frequency components of sound via an active electromotile response. Three types of neural elements are found within the cochlea. Rosenthal's canal contains the spiral ganglion with its afferent auditory neurons. Type I spiral ganglion cells compose ~90% of the spiral ganglion neurons and innervate the inner hair cells, with multiple neurons projecting to and innervating a single inner hair cell. Type II spiral ganglion cells make up 10% of the spiral ganglion's neuronal population; each type II neuron innervates multiple outer hair cells. The endolymphatic sac connects to the endolymph-filled portion of the inner ear via the endolymphatic duct. The sac is a complex structure that lies intercalated within the dura of the posterior fossa. An imaginary line drawn through the horizontal semicircular canal and the outline of the posterior semicircular canal approximates the position of the sac. The cochlear aqueduct connects the scala tympani via an opening on the petrous pyramid in the posterior fossa. In most humans this duct has been obliterated by connective tissue, but it forms a potential communication for bacteria to and/or from the labyrinth to the cerebrospinal fluid of the cranial cavity.

Three nerves innervate the sensory receptors of the inner ear: the cochlear nerve and the superior and inferior divisions of the vestibular nerve (see **Fig. 22–3**). The superior vestibular nerve supplies part of the saccule (via Voit's nerve), the utricle, and the superior and lateral semicircular canal ampullae. The inferior vestibular nerve innervates the posterior canal ampulla and the remainder of the saccule.

VASCULAR SUPPLY

The blood supply of the pinna is supplied by the posterior auricular artery (external carotid), the anterior auricular artery (superficial temporal), and a branch of the occipital artery. Venous drainage is supplied by the corresponding veins as well as by the mastoid emissary vein. Lymphatic drainage is to the anterior auricular nodes and then to the superficial parotid nodes anteriorly, or to the retroauricular nodes, and from there to the upper deep cervical nodes posteriorly or inferiorly. Within the temporal bone the internal carotid crosses under the eustachian tube and then turns upward and inward anterior to the cochlea. The major venous structure within the temporal bone is the sigmoid sinus. The sigmoid sinus runs from the lateral sinus to the jugular bulb that then becomes the jugular vein. The position of the jugular bulb is variable, and it may be dehiscent in the floor of the middle ear. The superior petrosal sinus connects the transverse sinus with the cavernous sinus. The inferior petrosal sinus connects the jugular bulb with the cavernous sinus. The middle ear is supplied by several different arteries derived from both the internal and external carotid as well as the basilar arteries. Vascular perfusion of the inner ear is supplied by the labyrinthine artery that is a branch of the anteroinferior cerebellar artery (AICA).

NERVES RUNNING THROUGH THE TEMPORAL BONE

The anatomy of the facial nerve is presented in detail in Chapter 35. However the facial nerve deserves a special comment in this chapter because of its complicated route through the temporal bone. The knowledge of its anatomy is a prerequisite for any surgery of the middle ear and the mastoid. The facial nerve enters the temporal bone by the internal auditory canal (IAC). Within the IAC the facial nerve is superior to the cochlear nerve and anterior to the superior vestibular nerve. At the lateral end of the IAC, the nerve passes superior to a ridge called the transverse crest and anterior to a wedge called Bill's bar. Lateral to this is the labyrinthine segment of the facial nerve. The nerve travels slightly anterior between the basal turn of the cochlea and the ampulla of the superior semicircular canal. The geniculate ganglion lies at the lateral end of the labyrinthine segment. Here the facial nerve turns posterior (first genu) to become the horizontal or tympanic segment, and the greater petrosal nerve exits the geniculate ganglion anteriorly and exits the temporal bone through the facial hiatus. The main trunk of the facial nerve passes along the medial wall of the tympanic cavity superior to the oval window niche and then bends around the oval window (second genu) to travel in an inferior direction as the vertical or mastoid segment begins. The facial nerve gives off

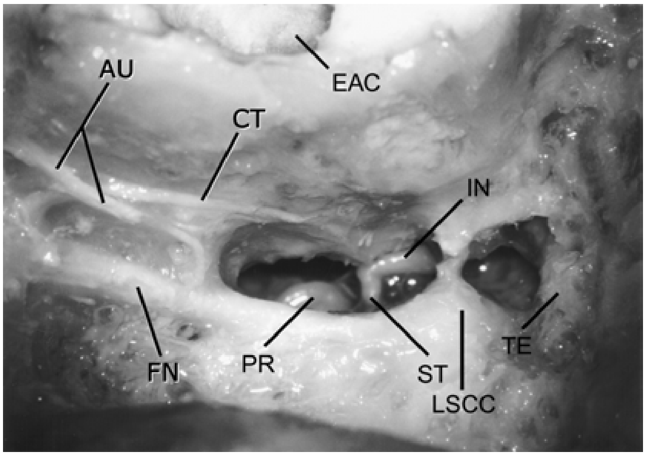


Figure 22–6 *A temporal bone after dissection of facial recess. AU, auricular branch of facial nerve; CT, chorda tympani; EAC, external auditory canal; FN, facial nerve; IN, incus; LSCC, lateral semicircular canal; PR, promontory; ST, stapes; TE, tegmen.*

three branches during its descent through the mastoid cavity: (1) the nerve to the stapedius muscle, (2) the sensory auricular branch of the facial nerve that innervates the EAC, and (3) the chorda tympani. These last two branches of the facial nerve are useful anatomical landmarks during surgery. Transmastoid surgical procedures often require a progressive thinning of the posterior wall of the EAC to identify the descending segment of the facial nerve and to access middle ear from the mastoid through the facial recess. The facial recess makes up the space between the chorda tympani, the main trunk of the facial nerve, and the incus (**Fig. 22–6**).

Jacobson's nerve or the inferior tympanic nerve is a branch of CN IX that enters the middle ear through the inferior tympanic canaliculus. This nerve then forms the tympanic plexus and may arborize to the promontory, where it may be submucosal, located within a sulcus on the promontory, or within a complete canal. The branches reunite if they have arborized close to the cochleariform process forming the lesser petrosal nerve; this nerve then passes in a canal beneath the tensor tympani muscle and into the middle cranial fossa, where it is medial to the greater superficial petrosal nerve.

Arnold's nerve is a composite structure, formed by the auricular branch of the vagus nerve, the glossopharyngeal nerve, and the facial nerve. The auricular branch of the vagus nerve courses from the superior ganglion of the vagus nerve and passes posteriorly over the dome of the jugular bulb to enter the anterior aspect of the fallopian canal just proximal to the chorda tympani nerve. At this point the sensory contribution from the facial nerve fuses with the sensory contribution of the vagus nerve to form a single nerve, which exits the sheath of the main trunk of the facial nerve as the auricular branch or exits the temporal bone as a single auricular branch of the facial nerve from stylomastoid foramen providing sensation to the EAC and the auricle. Glomus tumors arise from the collection of chemoreceptor tissue, which is found in high concentration associated with the parasympathetic fibers carried in CN IX and X. Similar collections of chemoreceptor tissue (i.e., glomus bodies) are found along the course of Arnold's and Jacobson's nerves in the tympanic canaliculus, over the promontory, in the retrofacial air cell tracts, and around the geniculate ganglion.

EPONYMS AND ANATOMICAL PEARLS

- Citelli's angle = sinodural angle.
- Dorello's canal: The VI nerve canal running between the tip of the petrous bone and the sphenoid bone. This is the anatomical basis for the triad of VI palsy, pain, and the draining of the ear observed in infections of the petrous apex.
- Huguier's canal: The exit point of the chorda tympani from the middle ear.
- Huschke's foramen: Found in the anteroinferior external auditory canal and connects to the preauricular area
- Hyrtl's fissure: An embryological connection between the hypotympanum and the subarachnoid space. It is usually obliterated in adults.
- Meckel's cave: A bony depression housing the CNV ganglion.
- Santorini's fissure: Lymphatics found in anteroinferior aspect of the external canal that allow for the spread of disease from the ear canal into the preauricular area and the skull base.
- Trautmann's triangle: The area between the sigmoid sinus, superior petrosal sinus, and lateral canal.

Suggested Readings

Donaldson JA, Duckert LG, Lambert PM, Rubel EW. Surgical Anatomy of the Temporal Bone. 4th ed. New York: Raven Press; 1992

Proctor B. Surgical Anatomy of the Ear and Temporal Bone. New York: Thieme Medical Publishers; 1989

Schuknecht HR, Gulya AJ. Anatomy of the Temporal Bone with Surgical Implications. Philadelphia: Lea & Febiger; 1986

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. The chorda tympani nerve is a branch of which cranial nerve?
   A. IX
   B. X
   C. VIII
   D. VII
2. The tensor tympani muscle is innervated by which cranial nerve?
   A. X
   B. VII
   C. V
   D. IX
3. The muscle responsible for eustachian tube closure is the
   A. Tensor tympani
   B. Tensor vela palatini
   C. Salpingopharyngeal
   D. None of the above
4. Gradenigo's syndrome is due to the anatomical relationships of
   A. The petrous apex to the CN VI nerve
   B. The mastoid to the sphenoid bone
   C. The facial nerve to the middle ear
   D. The cavernous sinus to the carotid

Chapter 23

HISTOLOGY AND HISTOPATHOLOGY OF THE TEMPORAL BONE

Joseph B. Nadol, Jr.

A working knowledge of the normal microscopic anatomy of the human temporal bone and the histopathology of disorders that affect hearing, balance, and facial nerve function is essential for practicing otolaryngologists as well as subspecialists in the area of otology and neurotology. This knowledge is helpful as a basis for understanding disease processes that will be encountered in the clinical setting, to evaluate the efficacy of medical management of disorders of the ear, and finally in perfecting or modifying surgical technique. The purpose of this chapter is to introduce the trainee to conventional light microscopic anatomy and to provide examples of common histopathology in a variety of disease categories. Although light microscopy of the human temporal bone has been done for over 100 years, progress in understanding the pathogenesis and treatment of otologic disorders requires renewed attention to this form of scientific inquiry. For example, cochlear implantation for the rehabilitation of the profoundly deaf using multichannel implants has been done on a routine basis for ~20 years. However, the biological consequences of implanting an electrode array in the human can most directly be answered by study of postmortem specimens from individuals who in life had undergone implantation. Furthermore, the current revolution in molecular genetics and molecular biology requires renewed attention to the collection and study of human specimens for several reasons. For example, much of what we know about the molecular genetics of hearing loss is based on homologous mutants or genetically engineered deletions in the mouse. However, comparison of the histopathology in these animal models with human disease is essential to verify the validity of the model. Furthermore, in addition to light microscopy, human specimens may be studied by a variety of techniques, including electron microscopy immunohistochemistry and retrieval and amplification of nucleic acid sequences, both from archivally collected and newly acquired temporal bone.

The purpose of this chapter will be to introduce the reader to the normal light microscopic anatomy of the human temporal bone and to provide a representative example of the pathology of the ear. For a more complete reference, the reader is referred to the classic text of Schuknecht (1993).

TECHNIQUES FOR RETRIEVAL AND STUDY OF THE HUMAN TEMPORAL BONE

The details concerning retrieval of temporal bones have been published elsewhere. The National Temporal Bone Hearing and Balance Pathology Resource Registry of the National Institute on Deafness and Other Communication Disorders provides an excellent resource for retrieval and study techniques, a national database of currently available temporal bone sections, and recent research publications in the area of temporal bone histopathology (Merchant et al, 1993). The registry may be accessed by a 24-hour telephone number (800–822–1327) or by e-mail (tbregistry@meei.harvard.edu).

PLANE OF SECTION

Although human temporal bone specimens may be sectioned in any anatomical plane, as depicted in the following examples, the most conventional plane of section is horizontal or axial, which corresponds to the axial plane in computed tomographic (CT) scanning. As in the following examples, the tissue has been embedded in celloidin, sectioned at 20 μ, and stained with hemotoxylin and eosin.

HISTOLOGY OF THE NORMAL TEMPORAL BONE

The 14 photomicrographs depicted in **Figs. 23–1A–K** are representative examples of serial horizontal sections through the left temporal bone of a male who died at age 77. The series of photomicrographs begin superiorly, near the floor of the middle cranial fossa, and pass inferiorly to the cranial base.

HISTOPATHOLOGY

Developmental Defects

The following are examples of common phenotypes of developmental defects in the temporal bone. They include anomalies of the cochlear capsule (Mondini's deformity), anomaly of the membranous labyrinth (Scheibe's dysplasia), congenital aural atresia, and the persistent stapedial artery.

Mondini's Anomaly (Fig. 23–2)

This dysplasia is characterized by anomalies of the bony capsule of the cochlea and vestibule as well as dysplasia of the membranous labyrinth. The degree of dysplasia and functional abnormality can vary widely. The length of the cochlear duct is shorter than the normal 32 to 34 mm, and in general the vestibule is large, with anomalies of the semicircular canals, including variations in size and number of canals. Mondini's dysplasia may occur with no other recognized genetic or nongenetic abnormalities,

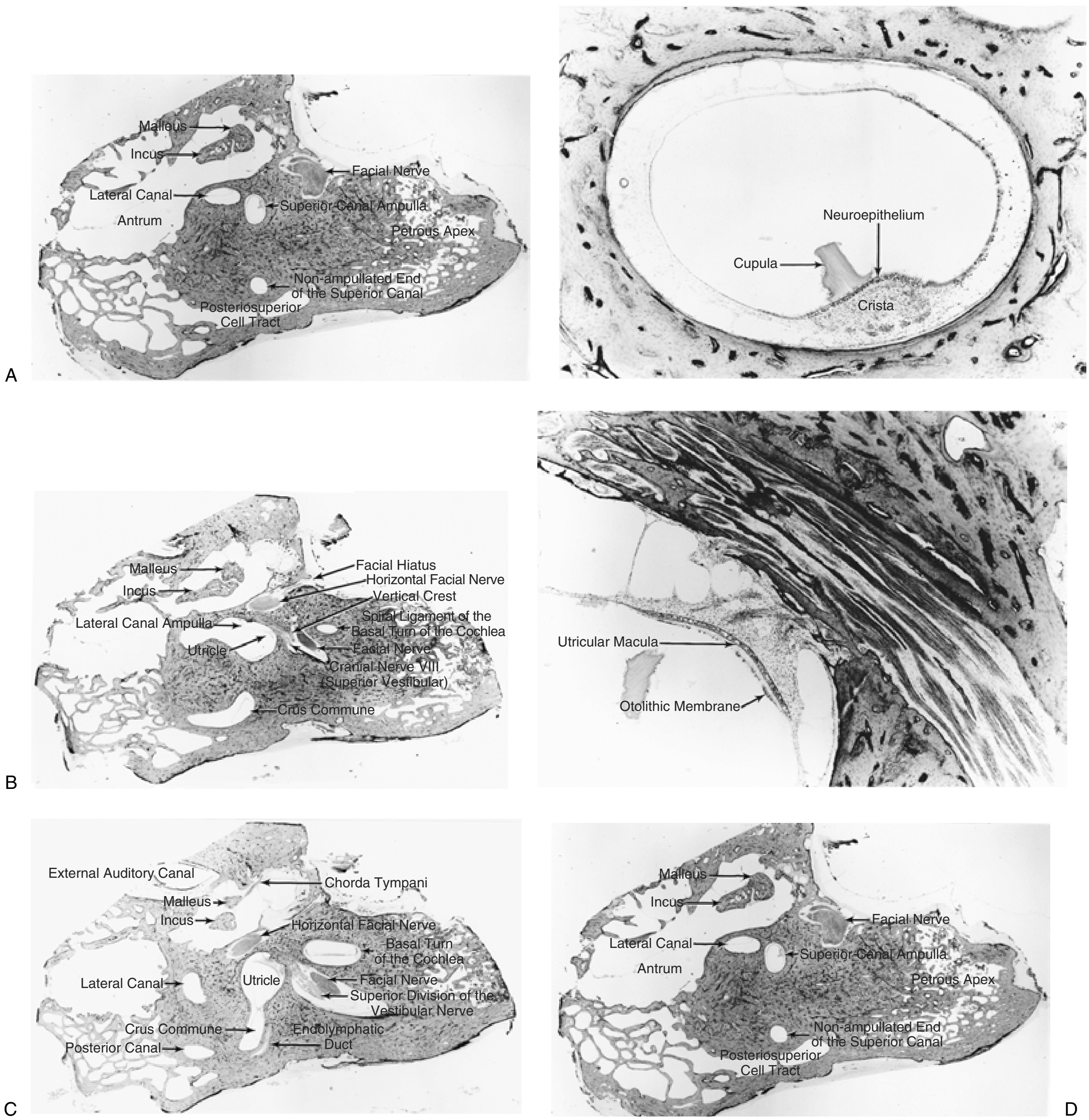


Figure 23–1 *Selected horizontal (axial) serial sections from a normal left adult human temporal bone.* ***(A, left)*** *Section at the level of the epitympanum, the head of the malleus and body of the incus, and through the arch of the superior semicircular canal, crista ampullaris of the superior semicircular canal, and geniculate ganglion. Pneumatized spaces include the antrum and the posterior superior cell tract (magnification X4).* ***(A, right)*** *Detail of the crista ampullaris of the superior semicircular canal (magnification X55).* ***(B, left)*** *Section at the level of the crista of the lateral semicircular canal, superior aspect of the internal auditory canal, and macula utriculi (magnification X4).* ***(B, right)*** *The macula utriculi and utricular division of the superior vestibular nerve (magnification X59).* ***(C)*** *Section at the level of the basal turn of the cochlea and the internal auditory canal (magnification X4).* ***(D)*** *Section at the level of the stapes footplate and tendon of the tensor tympani muscle (magnification X4).*

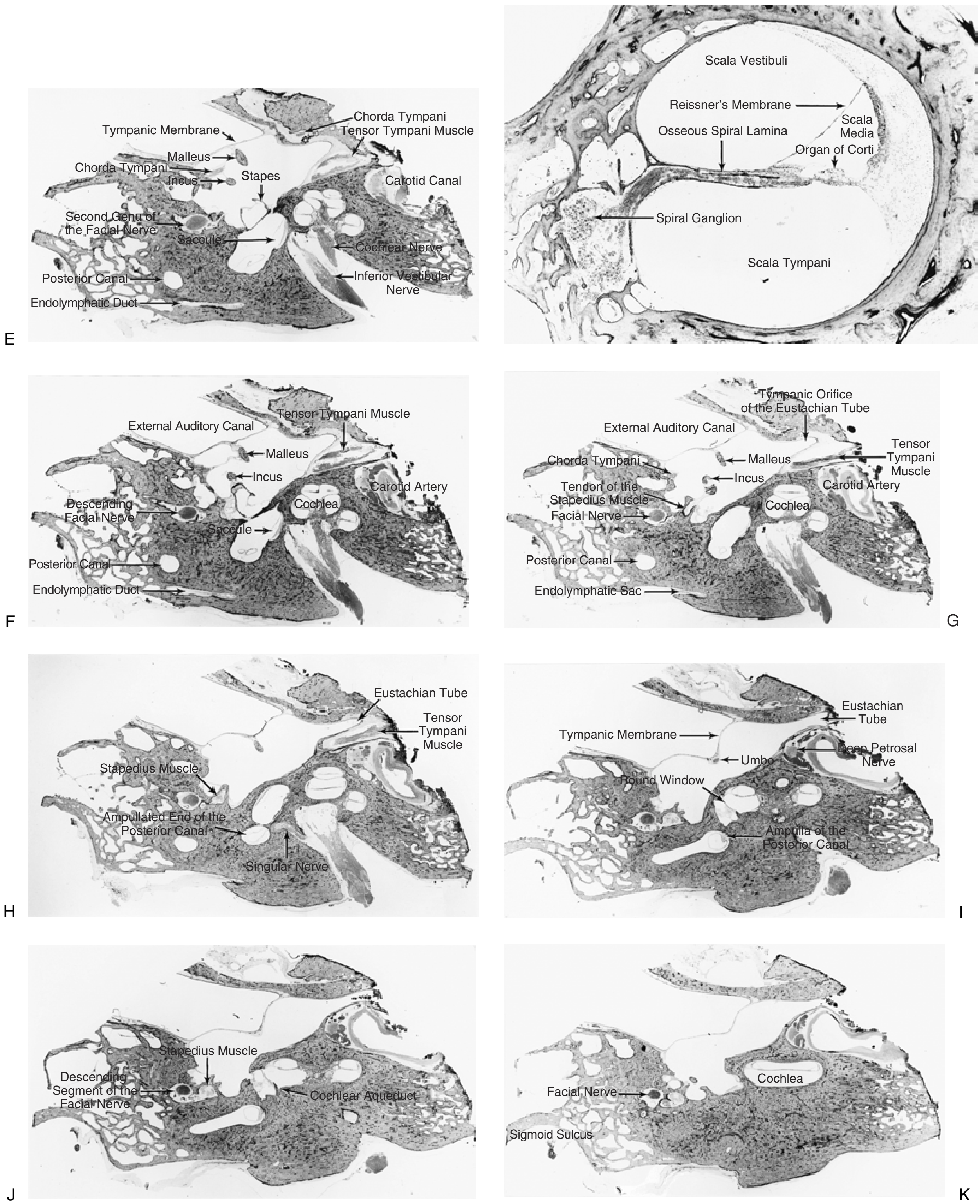


Figure 23–1 (Continued) ***(E, left)*** *Section through the modiolus of the cochlea (magnification X4).* ***(E, right)*** *Normal anatomy of the cochlea (magnification X57.8).* ***(F)*** *Section at the level of the macula sacculi (magnification X4).* ***(G)*** *Section at the level of the tendon of the stapedius muscle and descending segment of the facial nerve (magnification X4).* ***(H)*** *Section at the level of the ampullated end of the posterior semicircular canal and the singular nerve (magnification X4).* ***(I)*** *Section at the level of the round window membrane. The proximity of the eustachian tube to the petrous portion of the internal carotid artery is shown (magnification X4).* ***(J)*** *Section at the level of the cochlear aqueduct between the scala tympani of the basal turn and the subarachnoid space (magnification X4).* ***(K)*** *Section through the inferior aspect of the basal turn inferior to the round window (magnification X4).*

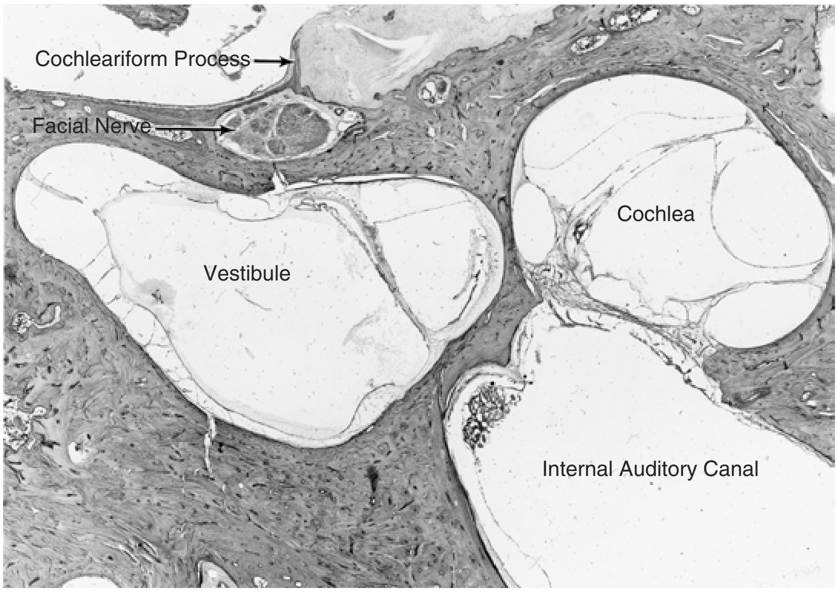


Figure 23–2 *Mondini's dysplasia of the left inner ear. This 85-year-old female was congenitally deaf, and severe Mondini's dysplasia was found in both inner ears. An enlarged and malformed vestibule and a cochlea consisting of a rounded compartment and malformed modiolus were seen (magnification X12).*

or it may be associated with known genetic disorders such as Klippel-Feil syndrome, Pendred's syndrome, and trisomies. Alternately, it may occur as the result of exposure to teratogenic insults in the fetus, such as diphenylhydantoin and isotretinoin. This dysplasia may be unilateral or bilateral and can be diagnosed on CT scan.

Scheibe's Dysplasia of the Membranous Labyrinth (Fig. 23–3)

In Scheibe's dysplasia, the bony capsule of the inner ear is normal. However, there is dysplastic development of the pars inferior (cochlea and saccule). Like Mondini's dysplasia, Scheibe's dysplasia is the phenotypic expression of a wide variety of genetically determined disorders of the inner ear. The dysgenesis may be seen in animals such as the Dalmatian dog.

Congenital Aural Atresia (Fig. 23–4)

This congenital anomaly is characterized by maldevelopment of the external auditory canal, resulting in either a narrowed canal or a true total atresia. It is commonly associated with anomalies of the auricle and both middle and inner ear. Congenital aural atresia may be seen as one manifestation of a genetically determined syndrome such as Treacher Collins, Crouzon's, branchio-oto-renal syndrome, or Goldenhar's syndrome, or in disorders with a questionable genetic basis such as the CHARGE syndrome (*c*oloboma of the eye, *h*eart anomaly, choanal *a*tresia, *r*etardation, and *g*enital and *e*ar anomalies), or as a consequence of intrauterine exposure to teratogenic substances.

Persistent Stapedial Artery (Fig. 23–5)

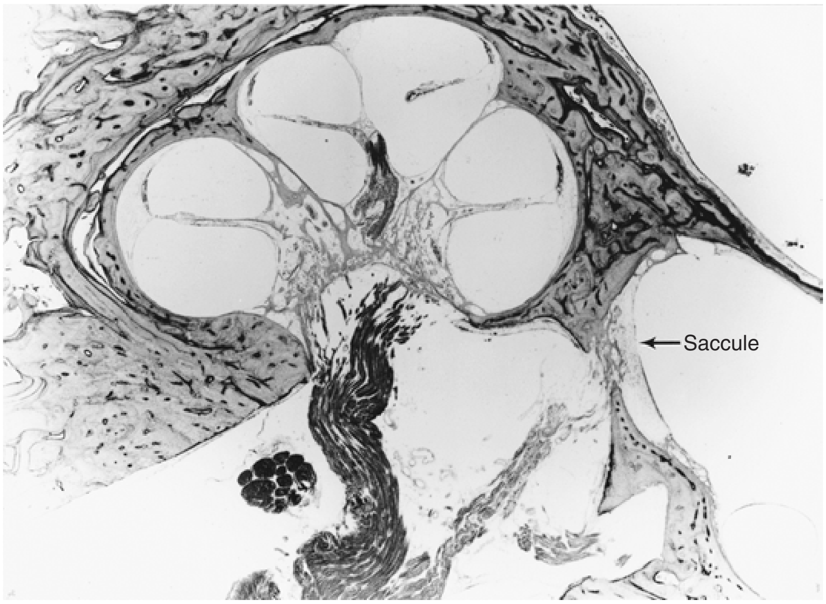
The embryonic circulation includes the stapedial branch of the hyoid artery. In the transition to the adult circulation, the stapedial artery atrophies, and the hyoid portion becomes the caroticotympanic arteries. However, in some cases the stapedial artery persists as a branch of the internal carotid system and supplies blood either to the distribution of the normal middle meningeal artery or to the distribution of the superorbital, infraorbital, and mandibular arteries. A persistent stapedial artery can be encountered during middle ear surgery as it passes through the obturator foramen of the stapes. Because this vessel may substitute for part of the intracranial blood supply, it should not be interrupted.

GENETICALLY DETERMINED DISORDERS OF THE INNER EAR

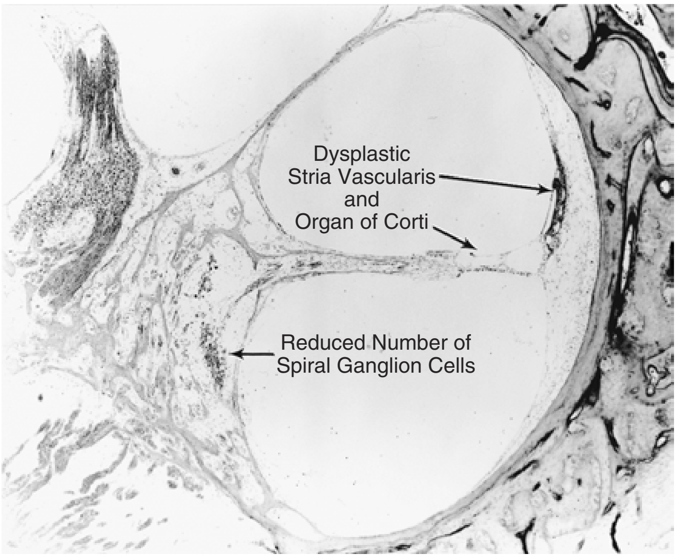
There is a burgeoning list of recognized genetic disorders of the inner ear. An excellent reference is the text by Gorlin et al (1995). This chapter presents only four representative examples: Usher's syndrome, Alport's syndrome, Waardenburg's syndrome, and the nonsyndromic dominantly inherited disorder DFNA-9.

Usher's Syndrome (Fig. 23–6)

Usher's syndrome is inherited as an autosomal recessive disorder and is characterized by sensorineural hearing



A



B

Figure 23–3 ***(A,B)*** *Scheibe's dysplasia of the right membranous labyrinth. This woman died at age 22 of renal failure. She was diagnosed at birth with Down syndrome and was congenitally deaf. The cochlea was somewhat short, consisting of two turns, and was 28 mm in length. The organ of Corti and stria vascularis were markedly dysplastic. There were no obvious hair cells. Although reduced in number, there were remaining spiral ganglion cells. The saccular macula was likewise dysplastic, whereas the cristae ampullares of the three semicircular canals and the macula utriculi were normal. Thus this patient demonstrated the classic histologic findings of cochleosaccular (Scheibe's) dysplasia of the membranous labyrinth (magnification of* ***A*** *X17;* ***B*** *X44).*

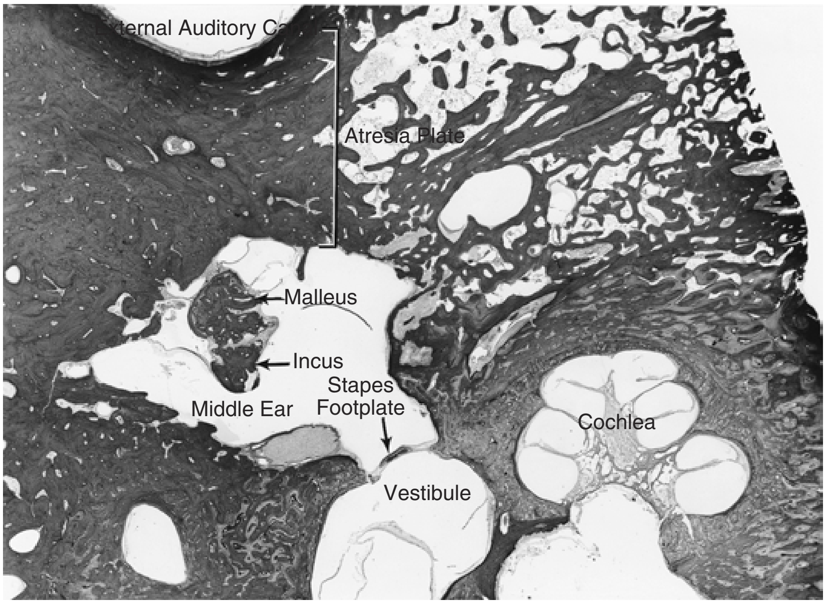


Figure 23–4 *Congenital left aural atresia in a 64-year-old male. There was a bony atresia plate, and the middle ear space was reduced in volume. The cochlea appeared normal (magnification X5.9).*

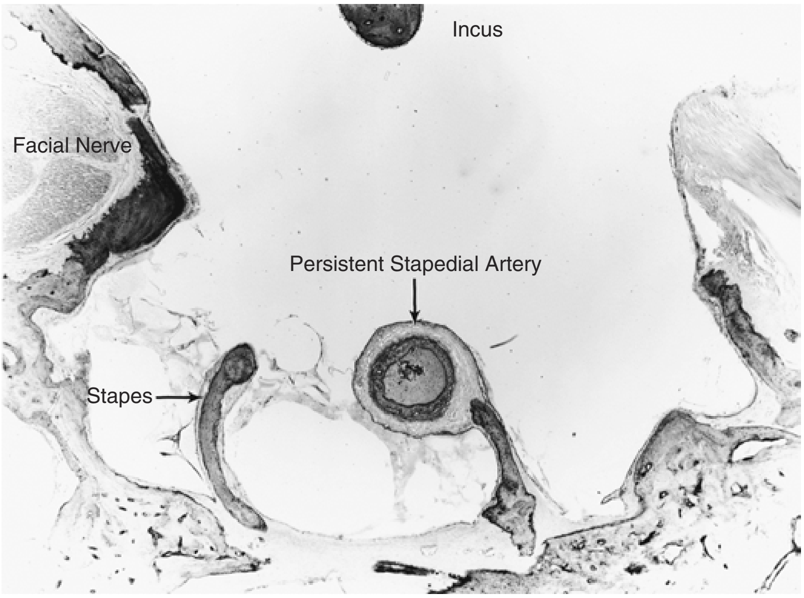
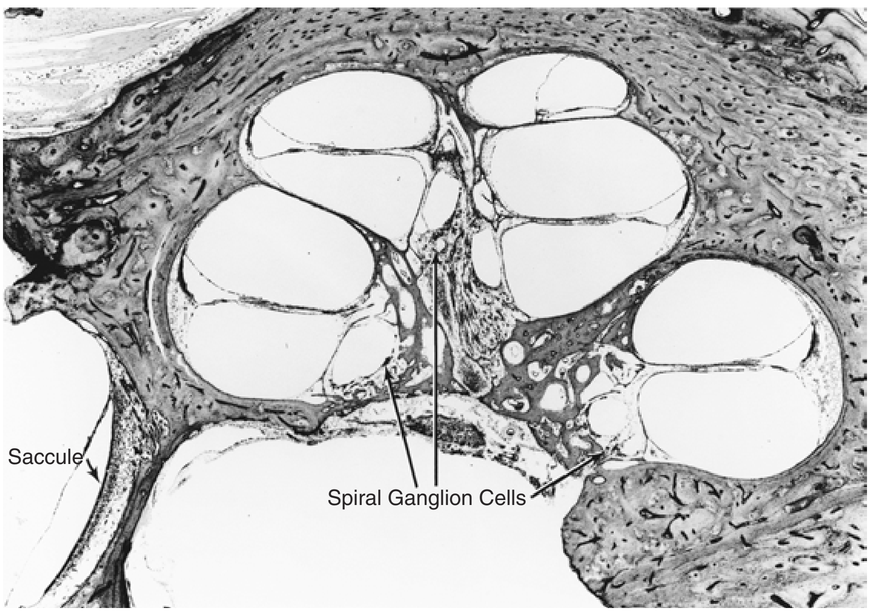


Figure 23–5 *Persistent stapedial artery in the left ear of an 84-year-old male. A large, persistent stapedial artery was seen in the obturator foramen of the stapes. On subsequent sections the artery branched within the fallopian canal, and the main trunk joined an anomalous internal carotid artery (magnification X26.9).*



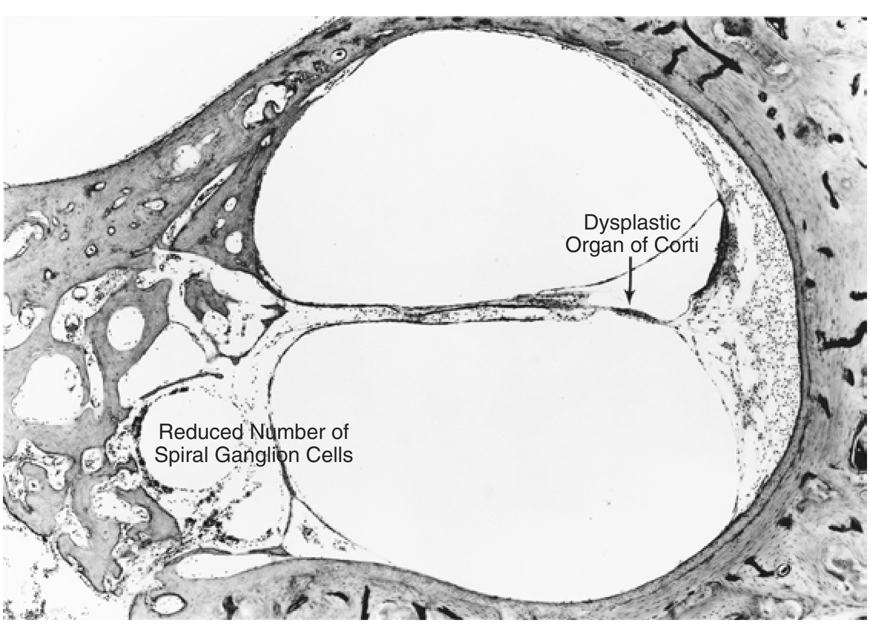


Figure 23–6 (A,B) *Usher's syndrome. This 84-year-old man was congenitally deaf and was diagnosed with Usher's syndrome type I. The bony labyrinth of this left ear was normal. There was marked degeneration or congenital dysplasia of the organ of Corti and spiral ganglion cells. In addition to loss of hair cells, there was marked atrophy or dysplasia of supporting elements of the organ of Corti. In contrast, the vestibular labyrinth, including the saccule, was normal (magnification of* ***A*** *X19.7;* ***B*** *X52).*

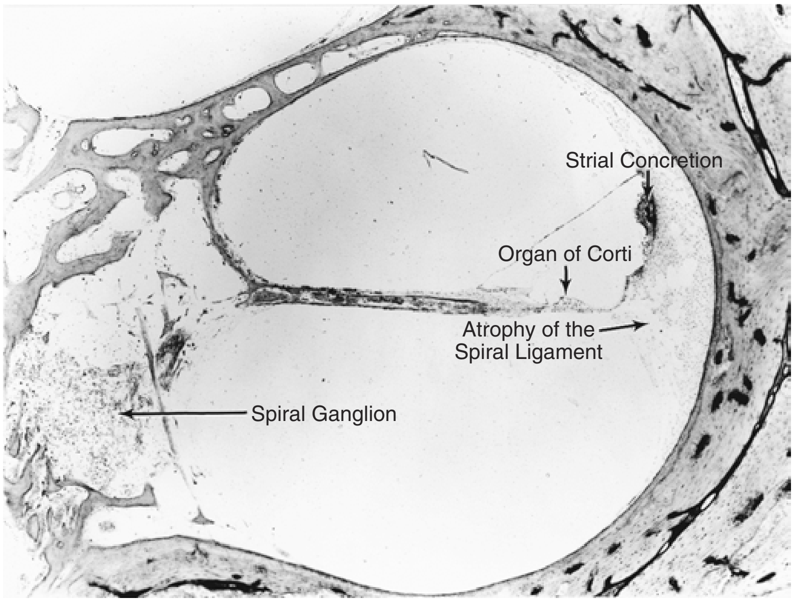


Figure 23–7 *Alport's syndrome. This 24-year-old man died from renal failure secondary to Alport's syndrome. Glomeronephritis was first diagnosed at age 8 and a sensorineural hearing loss at age 15. A progressive sensorineural loss was documented until the time of his death. In the organ of Corti of the left ear of the basal turn, there were strial concretions of extracellular material. There was loss of ~50% of spiral ganglion cells, whereas the hair cells appeared normal. In addition, there was cellular loss and atrophy of the spiral ligament (magnification X32).*

loss and retinitis pigmentosa. Three clinical forms have been recognized. Type I consists of severe to profound sensorineural loss, absent vestibular responses, and the development of retinitis pigmentosa by 10 years of age. Type II consists of a congenital moderate to severe sensorineural loss, normal vestibular responses, and the development of retinitis pigmentosa in the second to third decade. Type III consists of a progressive sensorineural loss with retinitis pigmentosa of variable onset. The type I and type II clinical variants are genetically heterogeneous (i.e., there are several genetic subtypes). To date, six different genetic linkages for type I Usher's syndrome have been identified. Only one such gene has been cloned, the so-called MYO7A at 11q13.5. The gene product is myosin 7A. It has been hypothesized that defects in this gene product interfere with the formation of stereocilia and perhaps movement of synaptic vesicles through the cytoplasm. The histopathologic correlate of Usher's syndrome includes degeneration of the organ of Corti and, in particular, the spiral ganglion.

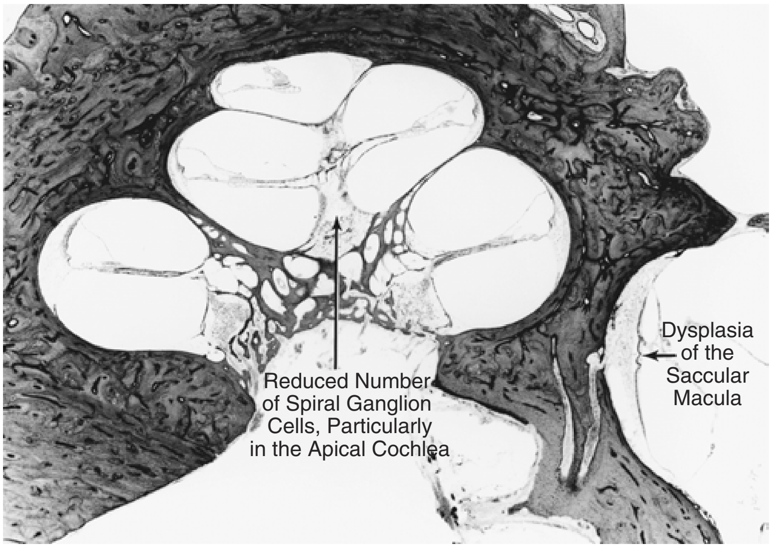
Alport's Syndrome (Fig. 23–7)

Alport's syndrome consists of progressive glomerulonephritis and sensorineural hearing loss. It is inherited as either X-linked or autosomal dominant, and at least six subtypes have been recognized. The X-linked forms are the most prevalent. Multiple mutations have been linked to the q22 region of chromosome X, and the gene has been cloned (COL4A5). The gene product is collagen 4A5, which is thought to play an important role in the structure of the spiral ligament and basilar membrane in the ear. The autosomal dominant form has been linked to human chromosome 2q36 to q37. Various mutations affect formation of collagen 4A3 and collagen 4A4, which are also thought to play a structural role in the basilar membrane of the inner ear. However, despite the implications of a structural dysfunction, the sensorineural hearing loss described in conjunction with Alport's syndrome is variable and progressive and worse in the high frequencies. Similarly, there are no typical histopathologic changes attributable to this syndrome, although degeneration of hair cells, cochlear neurons, and stria vascularis are commonly described.

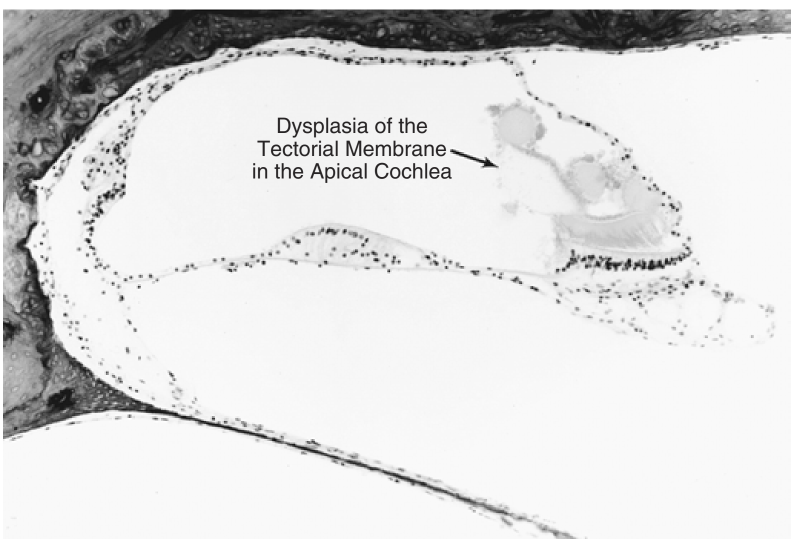
Waardenburg's Syndrome (Fig. 23–8)

Four clinical subtypes of Waardenburg's syndrome have been described:

> Type I: Inherited as an autosomal dominant disorder and characterized by the presence of dystopia canthorum. The genetic disorder has been linked to human chromosome 2q35 and is caused by mutations of the PAX3 gene, which is important for control of the transcription sequence. Therefore, it is assumed that this disorder is caused by abnormal migration of melanoblasts from the neural crest.
>
> Type II: Characterized by sensorineural hearing loss without dystopia canthorum. The genetic defect has been linked to 3p12.3 to p14.1 and is responsible for the transcription of a "zipper transcription factor." Dysfunction of this gene



A



B

Figure 23–8 (A,B) *Waardenburg's syndrome. This 76-year-old woman with Waardenburg's syndrome, including dystopia canthorum, white forelock, and heterochromia iridis, had a history of congenital hearing loss in her right ear only. Audiometry demonstrated a progressive low-frequency sensorineural hearing hoss with slight reduction in speech discrimination on the right. The hearing was normal on the left side. The organ of Corti was normal on the left, whereas on the right there was dysplasia of the tectorial membrane in the apical cochlea and reduced numbers of spiral ganglion cells, particularly in the apical half of the cochlea (magnification of **A** X18.1; **B** X140).*

product is felt to result in shortened survival of melanocytes.

Type III (Klein-Waardenburg syndrome): Consists of type I Waardenburg's and upper limb abnormalities. It has been described as a subtype of classic type I and is thought to be due to defects at chromosome location 2q35 and the *PAX3* gene.

Type IV (Waardenburg-Shah syndrome): Consisting of classic type II Waardenburg's and Hirschsprung's disease. It is inherited as an autosomal recessive. It is heterogeneous in at least three different gene locations, chromosome 13q22, 20q13.2 to 13.3, and 22q13. These genetic defects are thought to result in disorders of endothelial receptors that interfere with the development of the neural crest. In addition to sensorineural hearing loss and dystopia canthorum, other phenotypical abnormalities are described in some patients, including heterochromia iridis, white forelock, partial albinism, hypopigmentation of the ocular fundi, and broad and high nasal root. The cochlear histopathology in Waardenburg's syndrome includes degeneration of the organ of Corti and spiral ganglion and decreased numbers of melanocytes. An interesting feature of Waardenburg's syndrome is that the sensorineural hearing loss may be unilateral.

DFNA-9 (Fig. 23–9)

This is a dominantly inherited nonsyndrome sensorineural loss. The genetic defect has been cloned at human chromosome location 14q12 to q13. Mutations result in an abnormal COCH gene. This gene encodes

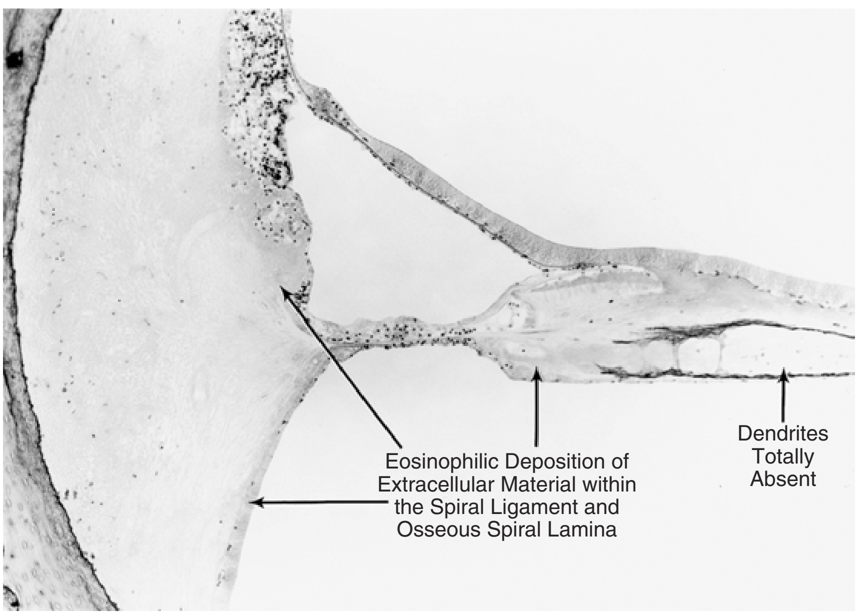


Figure 23–9 ***(A, B)****DFNA-9, right ear. This 59-year-old woman developed a progressive sensorineural hearing loss at age 21 that became profound by age 50. The genetic defect in this autosomally dominant inherited disorder has been cloned to chromosome 14. There was deposition of an eosinophilic extracellular material principally within the spiral ligament and osseous spiral lamina and also in the maculae and cristae. This finding appears to be specific to this disorder. Although the organ of Corti appeared normal, there were reduced numbers of spiral ganglion cells and total absence of dendrites of spiral ganglion cells within the osseous spiral lamina (magnification of* ***A*** *X24.1;* ***B*** *X144).*

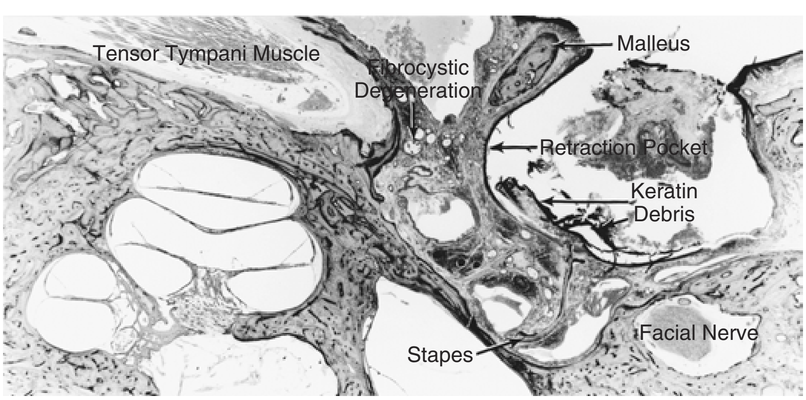


Figure 23–10 *Chronic otitis media, right ear. This 79-year-old woman had a perforation of the tympanic membrane and a cholesteatoma in the right middle ear. In this section at the level of the stapes, there was a retraction pocket of the tympanic membrane filled with keratin debris. In addition, the subjacent mucosa was markedly thickened by fibrous tissue and cyst formation (magnification X10.2).*

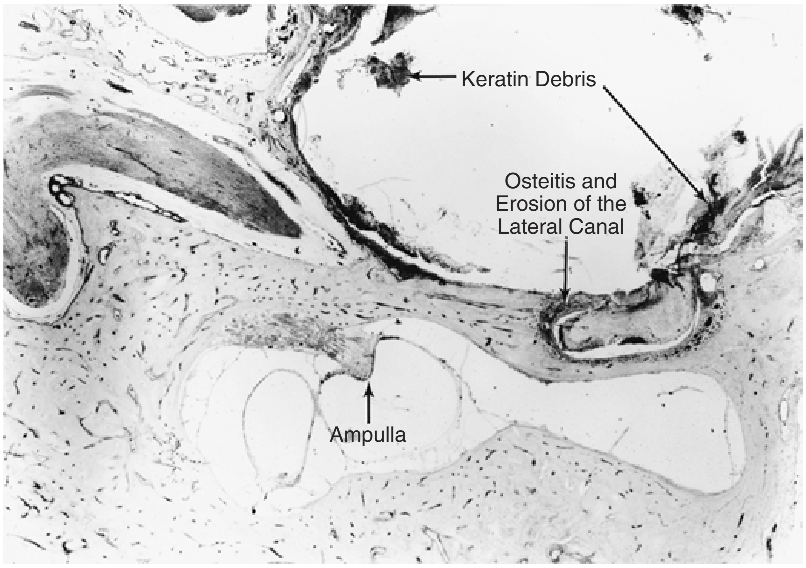


Figure 23–11 *Chronic otitis media. This 66-year-old man first developed otorrhea and hearing loss on the left at age 35. No surgery had been performed. During the last few weeks of life, he developed episodes of vertigo. The right middle ear and mastoid were normal. However, on the left there was a large posterior perforation with cholesteatoma. There was a resorptive osteitis over the lateral semicircular canal creating a fistula through the endosteum (magnification X16.1).*

the gene product cochlin, which appears to be exclusively expressed in the inner ear and thought to be involved in binding fibrillar collagens, glycoproteins, and proteoglycans. The cochlear histopathology is quite unique and consists of a deposition of extracellular material in a variety of locations in the inner ear, including the spiral ligament and osseous spiral lamina, and severe degeneration of dendritic processes of spiral ganglion cells.

INFECTIONS OF THE TEMPORAL BONE

Chronic Suppurative Otitis Media (Figs. 23–10 and 23–11)

The destructive potential of chronic suppurative otitis media in the human is best studied in the human temporal bone. Common changes include retraction pockets, cholesteatoma, resorption of ossicles, fibrocystic degeneration of the mucosa of the middle ear, and fistulas of the bony capsule of the inner ear.

Suppurative Labyrinthitis (Fig. 23–12)

Suppurative labyrinthitis may occur as an extension of an infectious process from the middle ear and mastoid such as acute or chronic otitis media or from the subarachnoid space in acute bacterial meningitis. In general, suppurative labyrinthitis, unless arrested very early in its course, results in profound sensory hearing loss due to destruction of the inner ear.

Petrous Apicitis (Fig. 23–13)

Petrous apicitis may occur as a complication of either acute or chronic otitis media. Gradenigo's syndrome, the classic triad of otorrhea, retro-orbital pain, and

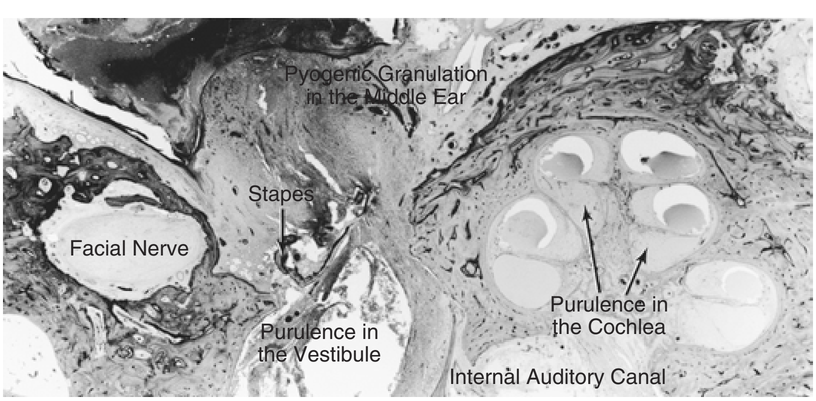


Figure 23–12 *Acute labyrinthitis. This 68-year-old woman had chronic otorrhea of the left ear despite previous mastoid surgery. Five weeks prior to death, she developed fever and vertigo. A diagnosis of otogenic meningitis was made. Despite treatment, she died of this disorder. The middle ear space was filled with pyogenic granulation that extended into the vestibule and cochlea. There was purulent exudate in the vestibule and cochlea (magnification X10.2).*

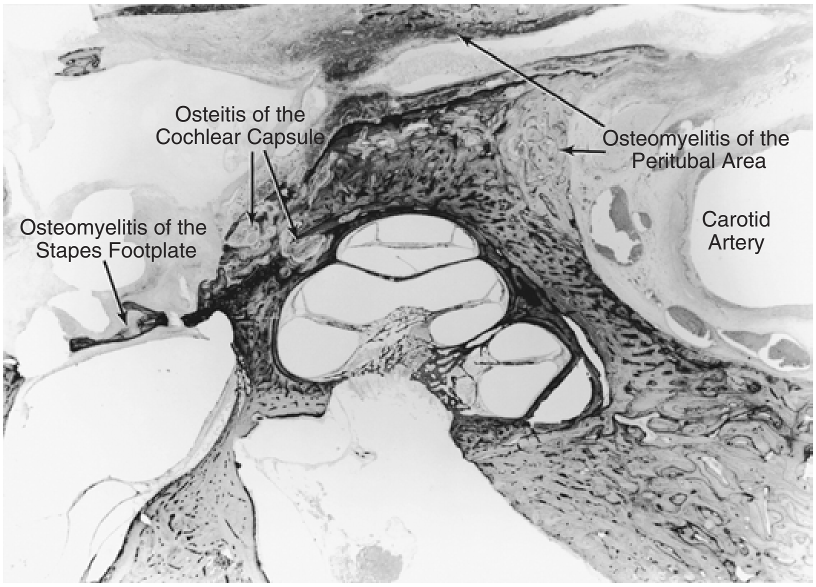


Figure 23–13 *Chronic petrous apicitis. This 70-year-old woman became profoundly deaf in her left ear a few months prior to death. Concomitantly, she developed ataxia. At age 67, a diagnosis of chronic lymphatic leukemia was made. On the left side, there was osteomyelitis of the temporal bone in the petrous apex and peritubal cell area adjacent to the carotid canal. There was also resorption of part of the stapes footplate and cochlear capsule (magnification X10).*

abducens paralysis, may be seen. Petrous apicitis may progress to subdural abscess, meningitis, or thrombosis or rupture of the carotid artery.

Syphilis (Fig. 23–14)

Syphilis may involve the temporal bone and inner ear in the secondary form of acquired syphilis associated with meningitis, late-latent acquired syphilis, tertiary or neurosyphilis, or as a consequence of congenital infection. The histopathology of syphilis of the inner ear includes a meningolabyrynthitis or a progressive osteitis of the otic capsule including microgummata. The hearing loss associated with syphilis is variable, ranging from sudden and complete to progressive or even fluctuant. It may mimic the auditory and vestibular symptoms of Meniere's disease.

Cytomegalic Inclusion Disease (Fig. 23–15)

Infection with the cytomegalovirus most commonly occurs in utero, or it may be acquired later in life, particularly in an immunologically impaired adult. In the congenital form, the presentation varies from death in utero to a relatively asymptomatic virema. In more severely affected but surviving individuals, neural involvement includes brain injury, blindness, and deafness.

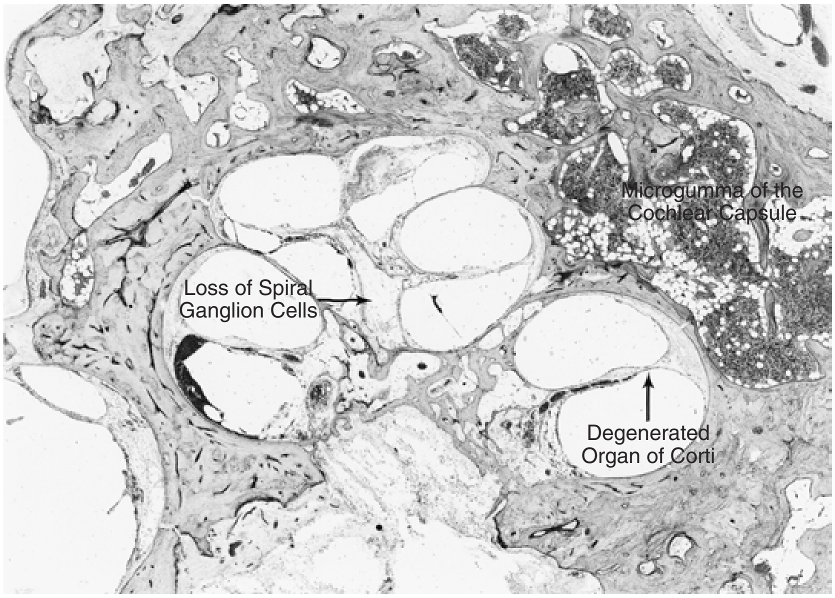


Figure 23–14 *Otosyphilis, left ear. This 91-year-old woman had congenital syphilis. She was profoundly deaf from at least early childhood. There were large areas of osteitis (microgummata) of the otic capsule with severe degeneration of the organ of Corti and marked loss of spiral ganglion cells (magnification X20).*

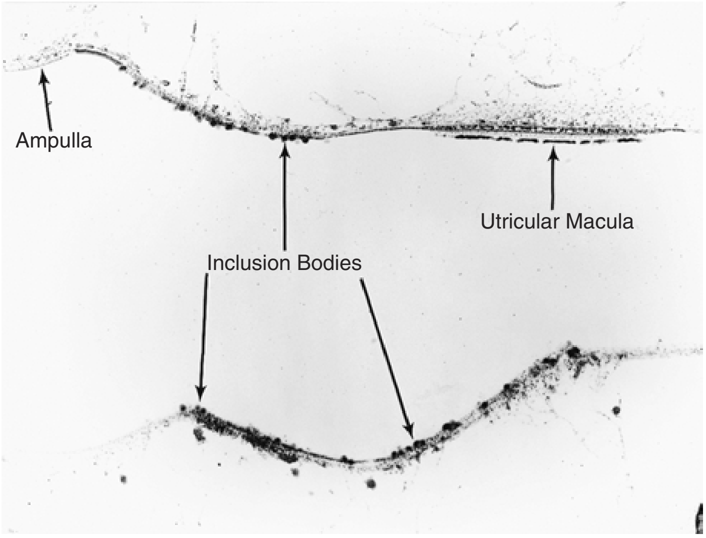


Figure 23–15 *Cytomegalic inclusion disease. This 11-day-old male died of hypotension and hemolysis secondary to disseminated congenital cytomegalic inclusion disease. In the left inner ear, there were multiple cytomegalic inclusions within the vestibular labyrinth (magnification X53.8).*

The histopathology of cytomegalic inclusion disease consists of a viral labyrinthitis resulting in degenerative changes of various tissues of the inner ear, including the stria vascularis and dark cell area of the vestibular labyrinth.

OTOTOXICITY

Aminoglycoside Ototoxicity (Fig. 23–16)

The aminoglycoside antibiotics may result in ototoxic degeneration of the inner ear. In general, ototoxicity is dose related. However, the genetic makeup of the individual may determine the severity of ototoxicity. For example, in the mitochondrial gene mutation 1555A-G, a maternally inherited susceptibility to aminoglycoside ototoxicity has been described.

Although progressive damage to the inner ear may be limited by cessation of most aminoglycosides, ototoxicity may progress for months following cessation of treatment with neomycin. The target tissue likewise varies among the various aminoglycoside antibiotics. For example, streptomycin and gentamicin first affect the vestibular hair cells, whereas dyhydrostreptomycin affects auditory hair cells preferentially.

Loop Diuretics (Fig. 23–17)

The diuretic action of the loop diuretics ethacrynic acid and furosemide is due to inhibition of resorption of sodium and water in the loop of Henle. In general, the loop diuretics may produce a transient reversible sensorineural loss and vestibular dysfunction with high intravenous doses. However, in renal failure, hearing loss may be permanent. In addition, there is evidence to suggest that aminoglycosides and furosemide are synergistic in their damage to the inner ear.

Cisplatin (Fig. 23–18)

Cis-platinum is a commonly used chemotherapeutic agent. It may produce a high-frequency sensorineural hearing loss that is dose related. Further, there is evidence of synergy with kanamycin and with cranial irradiation.

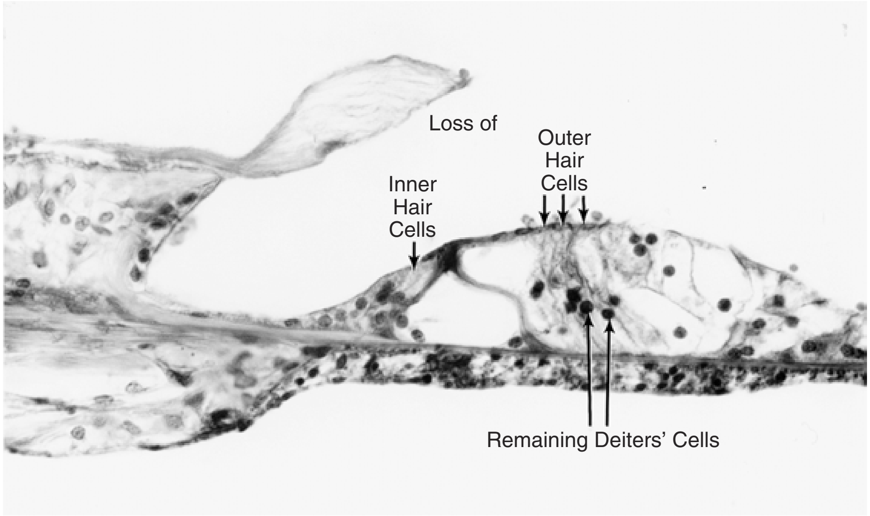
TRAUMA TO THE TEMPORAL BONE

Transverse Fracture of the Temporal Bone (Fig. 23–19)

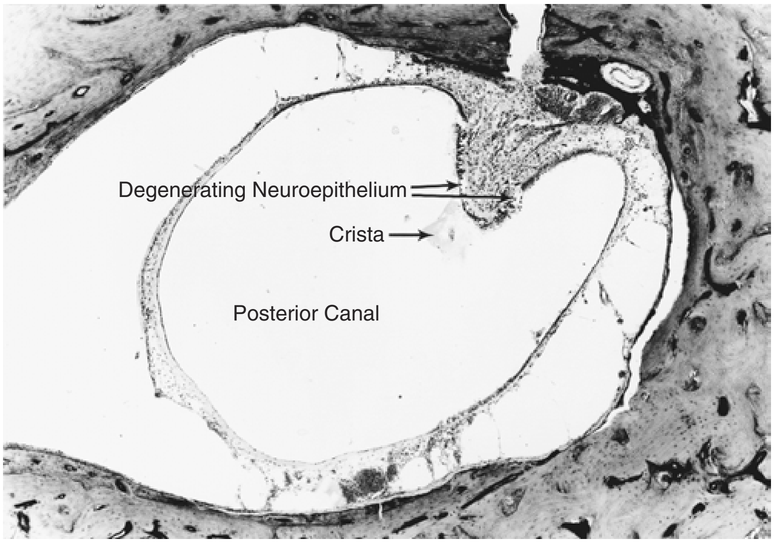
Trauma, particularly occipital trauma, may result in a fracture that is perpendicular to the long axis of the petrous pyramid (i.e., a transverse fracture of the temporal bone). Because the fracture may pass through the internal auditory canal and/or cochlea, injuries to the organ of Corti and neural elements within the internal auditory canal are common, resulting in sensorineural loss, vestibular disturbance, and/or facial paralysis.

Noise Trauma (Fig. 23–20)

Noise trauma to the ear may produce sensorineural hearing loss that is more severe in the high frequencies. This may be the result of repeated exposure to occupational noise, or it may be the result of a single exposure to extremely loud noise, such as to an explosion. There is some evidence that there may be genetic variations in



A



B

Figure 23–16 ***(A,B)*** *Gentamicin ototoxicity, left ear. This 51-year-old man was treated with 80 mg of gentamicin intravenously every 12 hours for* Escherichia coli *septicemia. During the period of antibiotic administration, profound bilateral hearing loss and renal insufficiency occurred. There was near total loss of both inner and outer hair cells in all turns of the cochlea (**A**; magnification X504) and degeneration of the neuroepithelium of the cristae, including the posterior canal (**B**; magnification X53.8).*

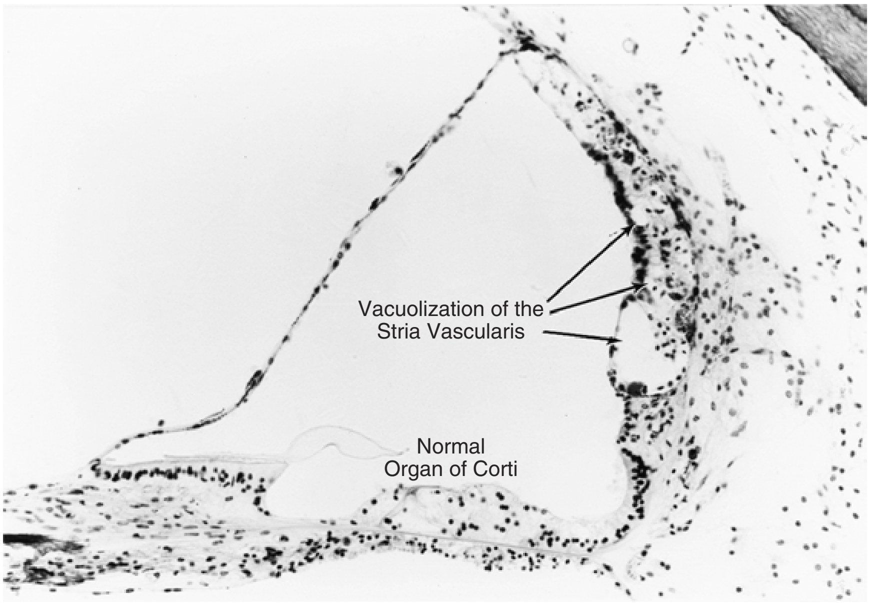


Figure 23–17 *Loop diuretic ototoxicity, left ear. This 65-year-old woman with carcinoma of the lung and arteriosclerotic cardiovascular disease received over 2 g of furosemide in the last 24 hours of life. There was marked extracellular vacuolization of the stria vascularis, particularly adjacent to the spiral prominence. The organ of Corti and auditory nerves were normal (magnification X200).*

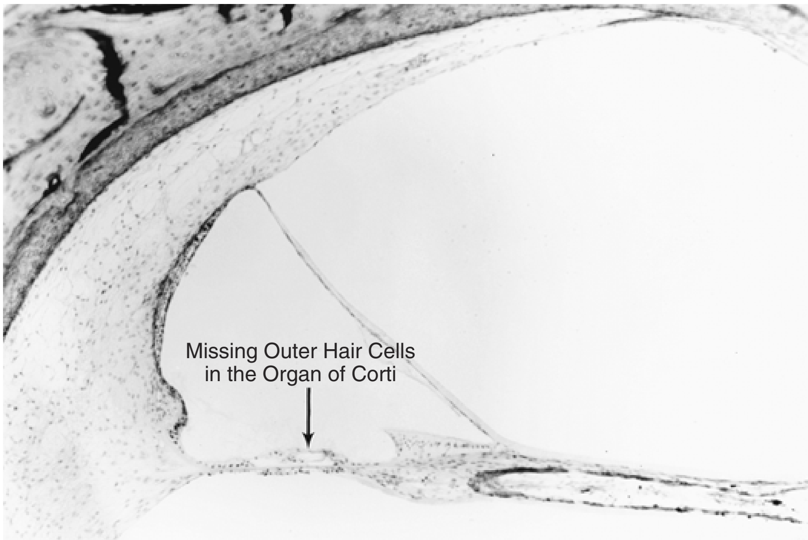


Figure 23–18 *Cisplatin ototoxicity, right ear. This 14-year-old boy was treated with irradiation and intrathecal methotrexate and cisplatin for meduloblastoma of the cerebellum. An audiogram 1 month following chemotherapy demonstrated a bilateral sensorineural loss, worse in the high frequencies, with 60% speech discrimination on the right. There was total loss of outer hair cells in the basal turn and secondary degeneration of cochlear neurons (magnification X116).*

sensitivity to a given noise level. In addition, the inner ear may be sensitized to noise by the administration of aminoglycosides or aspirin. Following noise trauma, the earliest changes occur in the basal turn, particularly among outer hair cells. With more severe or repeated injury, the inner hair cells may be damaged, with subsequent secondary degeneration of cochlear neurons.

VASCULAR DISORDERS

Vertebrobasilar Infarct (Fig. 23–21)

Blood to the inner ear is supplied by the labyrinthine artery, a branch of the anteroinferior cerebellar artery or the vertebrobasilar system. Hence, ischemia of the vertebrobasilar arteries may result in acute vestibular and auditory loss, most commonly associated with multiple other neurological symptoms such as hemiparesis, dysarthria, and diplopia.

Subarachnoid Hemorrhage (Fig. 23–22)

Although other central nervous system (CNS) signs are predominant, subarachnoid hemorrhage may produce at least transient auditory and vestibular symptoms. This may be due to extravasation of blood products into Rosenthal's canal and perilymphatic

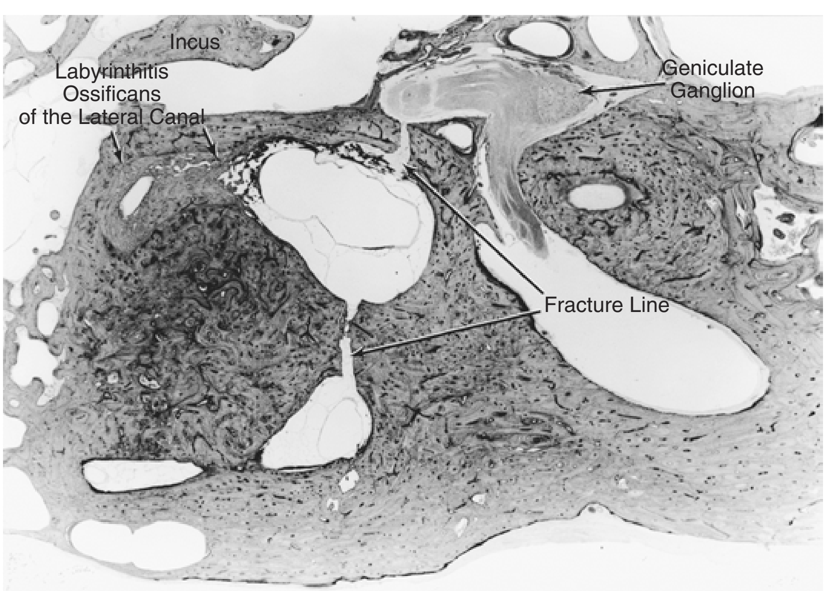


Figure 23–19 *Transverse fracture of the left temporal bone. This 59-year-old man became profoundly deaf in the left ear following a blow to the head 50 years before death. There was a fracture line that was largely unhealed and passed through the fallopian canal, vestibule, and crus commune. There was degeneration of the organ of Corti and spiral ganglion cells. There was also labyrinthitis ossificans (new bone formation) within the lumen of the lateral semicircular canal (magnification X7.1).*

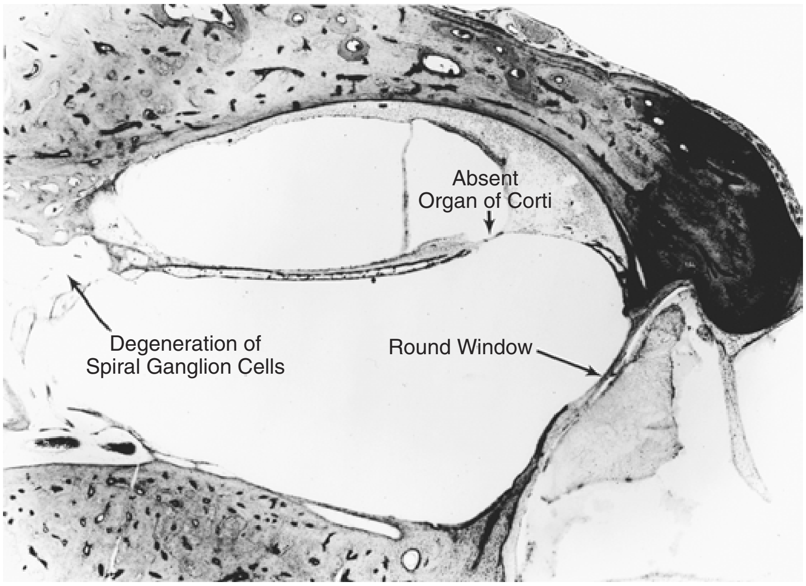
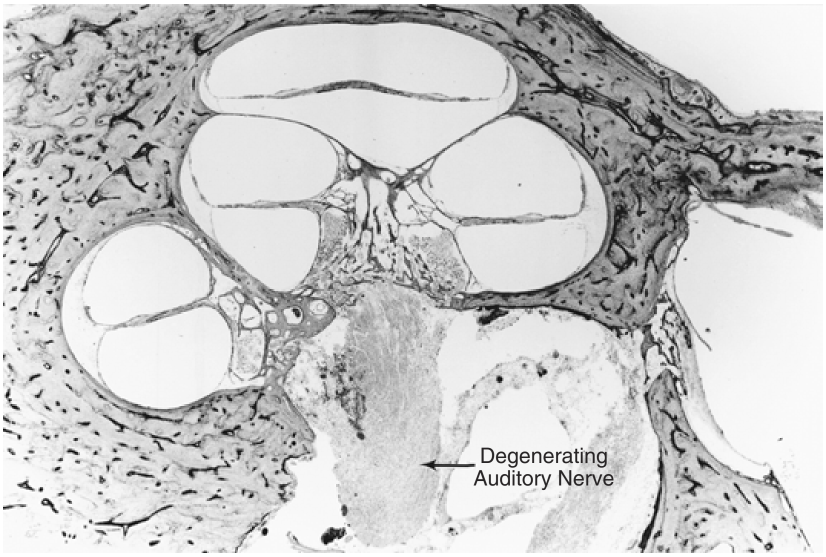
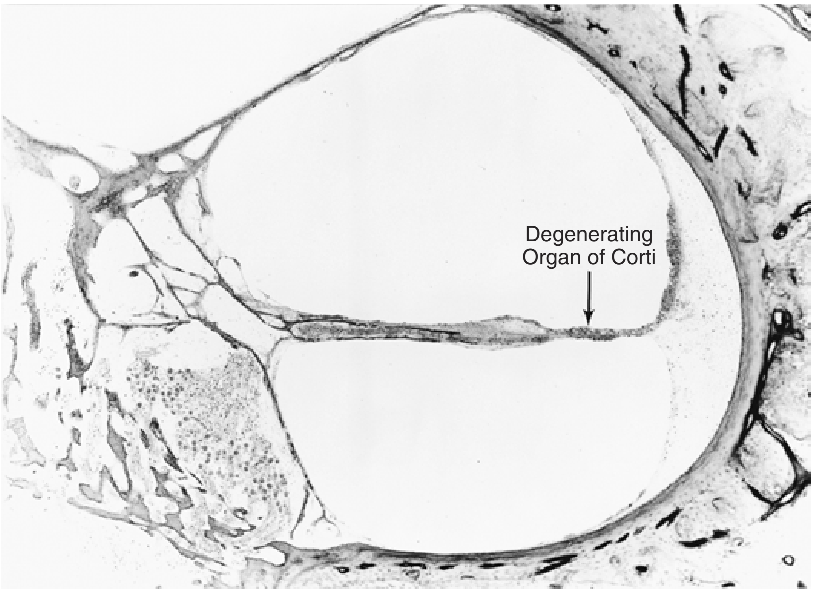


Figure 23–20 *Noise trauma. This 71-year-old retired machine worker had a documented high-frequency sensorineural loss in both ears, with 58% speech discrimination on the right. There was degeneration of the organ of Corti and spiral ganglion cells in the basal 15 mm of the right cochlea (magnification X31.5).*



A



B

Figure 23–21 (A,B) *Vertebrobasilar infarct. This 65-year-old man died as a consequence of brainstem infarction secondary to a right vertebrobasilar thrombosis. In the right ear there was severe degeneration of the membranous labyrinth, eighth nerve, and vestibular labyrinth. There was also loss of cellularity of the spiral ligament (magnification of **A** X17.7; **B** X53.5).*

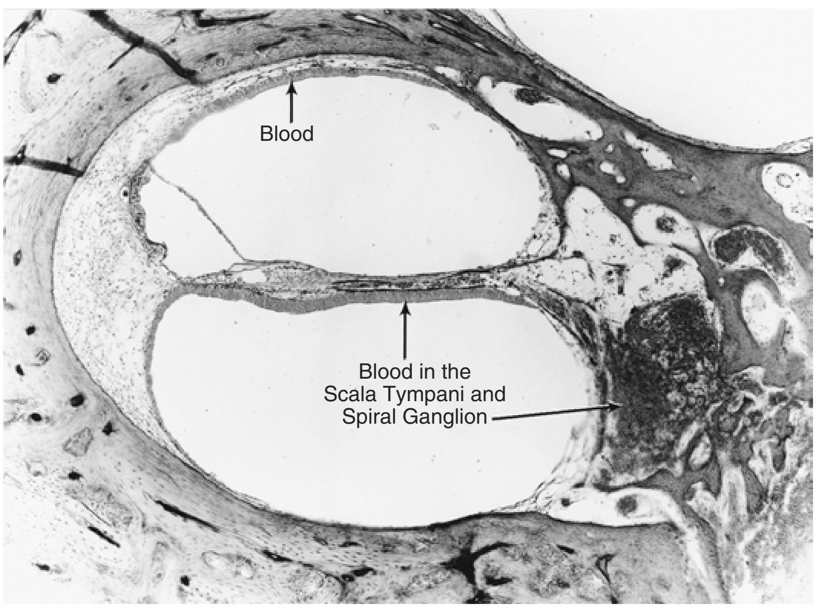


***Figure 23–22** **(A,B)** Subarachnoid hemorrhage, right ear. This 77-year-old woman died of subarachnoid hemorrhage originating in the left hemisphere. Prior to this illness, she was known to have a bilateral symmetrical flat 60 dB sensorineural hearing loss secondary to atrophy of the stria vascularis. There was fresh blood within the internal auditory canal that extended through the cribrose area into Rosenthal's canal, with blood products within the perilymphatic scalae (magnification of **A X**15.8; **B X**58.1).*

spaces of the inner ear via the internal auditory canal and cribrose area.

IMMUNE-MEDIATED SENSORINEURAL HEARING LOSS

Associated with Systemic Disease (Fig. 23–23)

Systemic immune disorders such as polyarteritis nodosa, Wegener's granulomatosis, temporal arteritis, lupus erythematosus, and Cogan's syndrome may produce concomitant audiovestibular dysfunction. Cogan's syndrome is characterized by an inflammatory disorder of the sclera and fluctuating auditory vestibular function, mimicking otosyphilis.

Without Systemic Disease (Fig. 23–24)

Immune-mediated sensorineural loss with no evidence of systemic involvement also may occur. The characteristic pattern is a rapidly progressive or fluctuating, usually bilateral, sensorineural loss, with decreased speech discrimination and vestibular disturbance. Although the erythrocyte sedimentation rate may be elevated in this disorder, the most specific laboratory test specificity (elevation of serum heat shock protein 70).

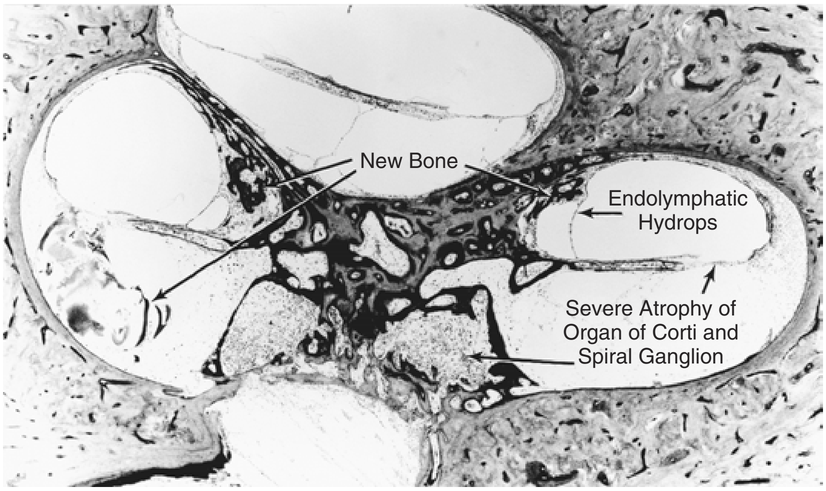
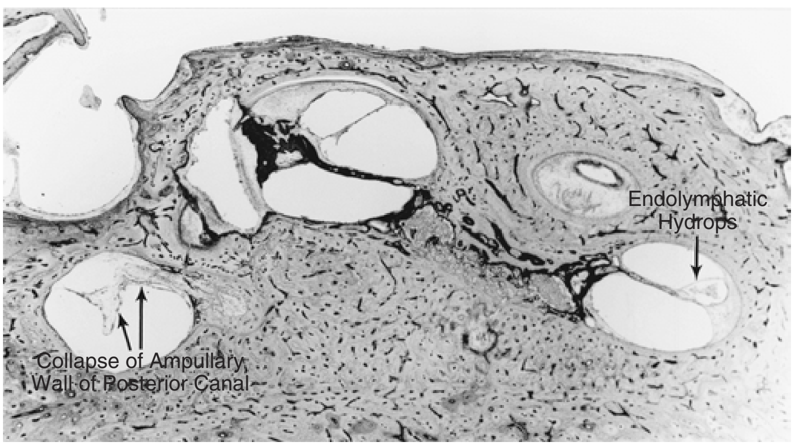


Figure 23–23 *Cogan's syndrome, left ear. This 62-year-old woman developed vertigo and progressive sensorineural hearing loss in both ears at age 31. She became profoundly deaf in both ears within 2 months of onset. One year later she developed uveitis and keratitis, and a diagnosis of Cogan's syndrome was made. There was severe atrophy of the organ of Corti and spiral ganglion. There was new bone within the membranous labyrinth and marked endolymphatic hydrops (magnification X26.6).*

A



B

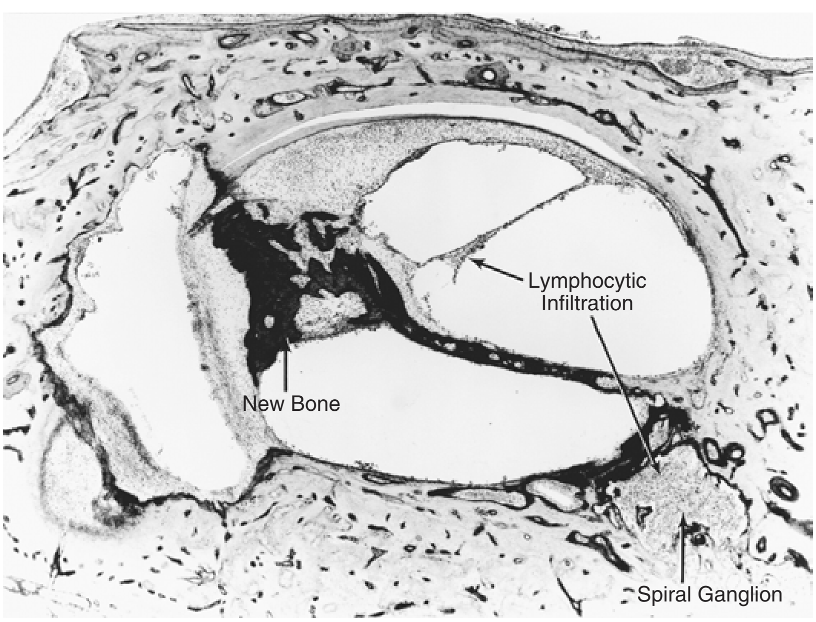


Figure 23–24 (A,B) *Immune-mediated inner ear disease, left ear. This 63-year-old man developed vertigo and ataxia at age 53. Over the subsequent years he had repeated episodes of vertigo and the onset of a progressive bilateral sensorineural hearing loss, worse in the left ear with reduction in speech discrimination. Electronystagmography demonstrated reduced vestibular response bilaterally. A tentative diagnosis of Meniere's disease was made. His hearing and balance function deteriorated, and at the time of death he was profoundly deaf in both ears. Histologic examination of the inner ear demonstrated diffuse inflammatory changes with lymphocytic infiltration and new bone formation within the cochlea. There was collapse of the walls of the posterior semicircular canal and severe degeneration of the auditory and vestibular neuroepithelium. There was hydrops of the cochlear duct. The histopathology is consistent with a chronic immune-mediated inflammatory disorder of the inner ear (magnification of **A** X11.8; **B** X31).*

Figure 23–25 *Otosclerosis. This 65-year-old woman first developed a mixed hearing loss in both ears at age 10. She had undergone bilateral stapedectomy later in life. Despite a transient improvement in hearing, there was a progressive loss of bone conduction, and prior to death she became profoundly deaf. There was an otosclerotic focus involving the right cochlear capsule to the level of the endosteum in all turns. There was subjacent degeneration of the spiral ligament and organ of Corti (magnification X15.5).*

DISORDERS OF BONE

OTOSCLEROSIS (FIGS. 23–25 AND 23–26)

Otosclerotic dysplasia of the cochlear capsule is extremely common. Approximately one in 10 individuals have histologic evidence of such involvement, and one in 100 may develop symptoms including conductive hearing loss due to stapediovestibular joint fixation, sensorineural hearing loss due to involvement of the endosteum of the cochlear capsule and subsequent degeneration of subjacent membranous labyrinth, and mild vestibular symptoms.

The etiology of otosclerosis is probably multifactorial. Certainly there is evidence for an autosomal dominant inheritance. In addition, there is evidence of infection of the dysplastic bone with measles virus. Otosclerosis is limited to the human temporal bone.

PAGET'S DISEASE OF BONE (FIG. 23–27)

This bony dysplasia may be due to paramyxovirus infection within osteoclasts. Involvement of the skull may include the temporal bone with deformity or

Figure 23–26 *Otosclerosis of the right anterior stapediovestibular joint. This 72-year-old woman had a bilateral mixed hearing loss. She had worn a hearing aid in her right ear for 20 years prior to death. There is a large otosclerotic lesion anterior to the stapes that caused fixation of the stapes. In this section there were various histologic stages of otosclerosis, including vascular basophilic active otosclerosis and more eosinophilic, less vascular sclerotic dysplastic bone (magnification X26.5).*

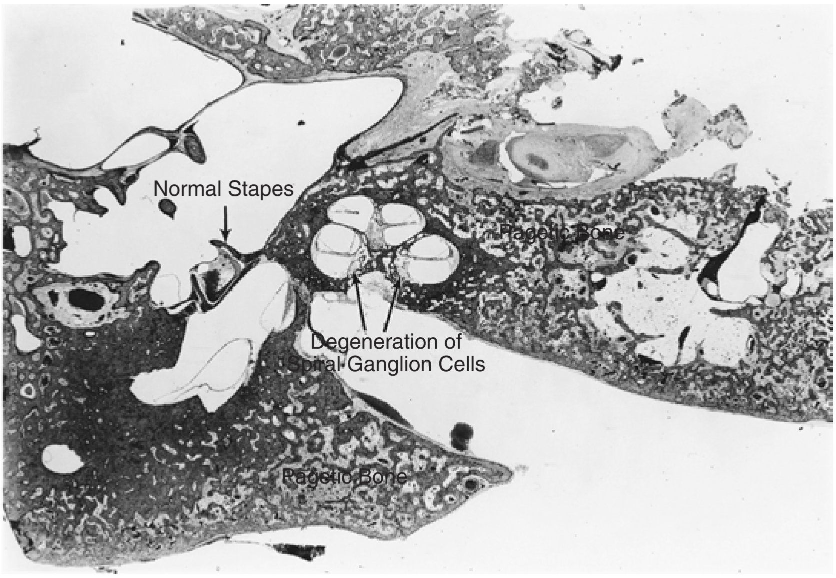


Figure 23–27 *Paget's disease of the temporal bone. This 85-year-old man developed a mixed hearing loss in both ears in his mid-70s. Histologic examination of the left ear demonstrated replacement of most of the temporal bone with pagetic dysplastic bone. There was no evidence of ossicular fixation; hence the conductive loss was unexplained. There was severe neural degeneration of auditory neurons throughout the cochlea (magnification X3.5).*

obliteration of the mastoid air cells. Likewise, compression of the internal auditory canal may occur. Only rarely are ossicles involved. Although commonly associated with a conductive hearing loss, this is not attributable to ossicular dysfunction. Rather, it probably represents an audiometric artifact related to abnormal sound conduction through the pagetic bone. Although sensorineural hearing loss is commonly described in pagetic patients, there is no consistent histopathologic correlate.

Osteogenesis Imperfecta (Fig. 23–28)

The classic syndrome of multiple fractures, blue sclera, and conductive deafness has several clinical variants based on severity of involvement and age of presentation. The disease demonstrates genetic heterogeneity resulting in a variety of mutations affecting the synthesis and assembly of collagen type I.

A conductive hearing loss has been described in types I, III, and IV and is caused by involvement of the ossicular chain, most commonly fixation of the stapediovestibular joint and/or a fracture of the long process of the incus.

AGING (PRESBYCUSIS)

There is a progressive increase in the prevalence of hearing loss concomitant with the aging process.

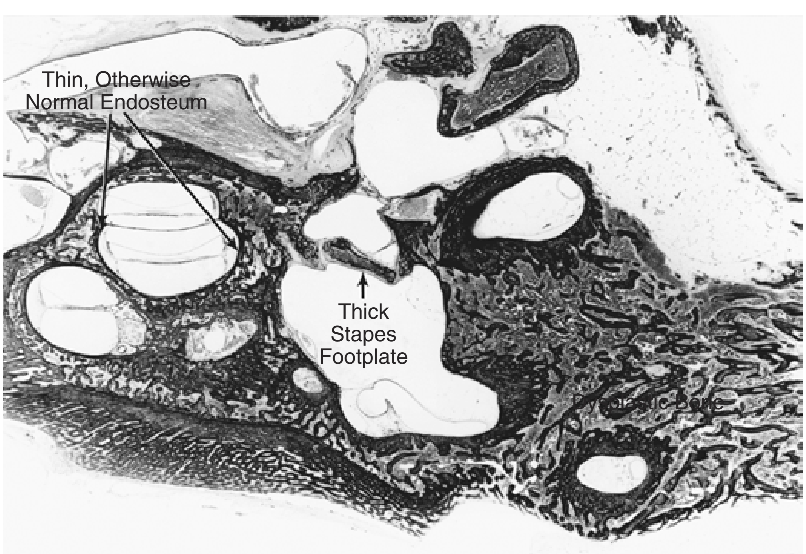
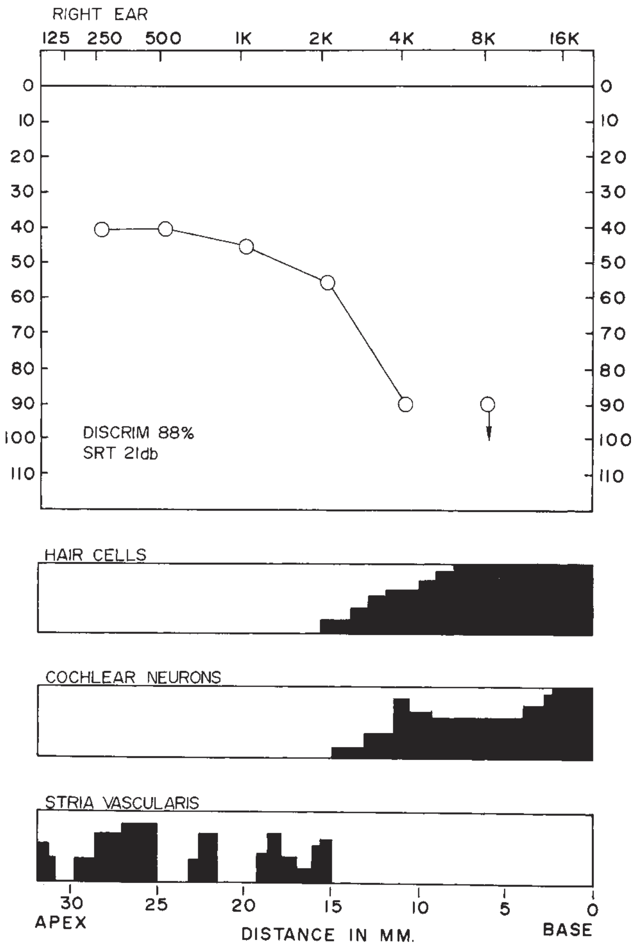
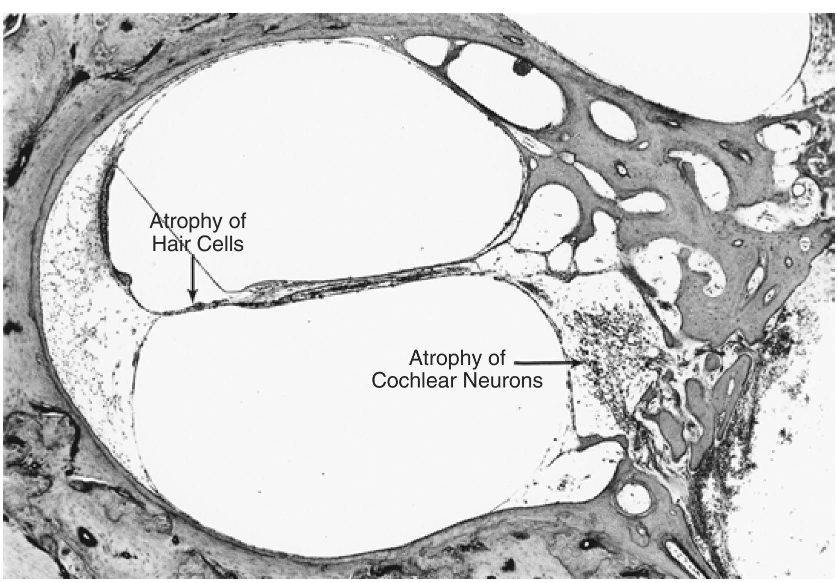


Figure 23–28 *Osteogenesis imperfecta, right ear. This newborn infant died within the first day of life secondary to severe osteogenesis imperfecta. There was dysplastic bone throughout the temporal bone, including the cochlea capsule and stapes footplate. Only the endosteal layer of bone appeared normal histologically. There was no evidence of ossicular fixation (magnification X8.9).*



A



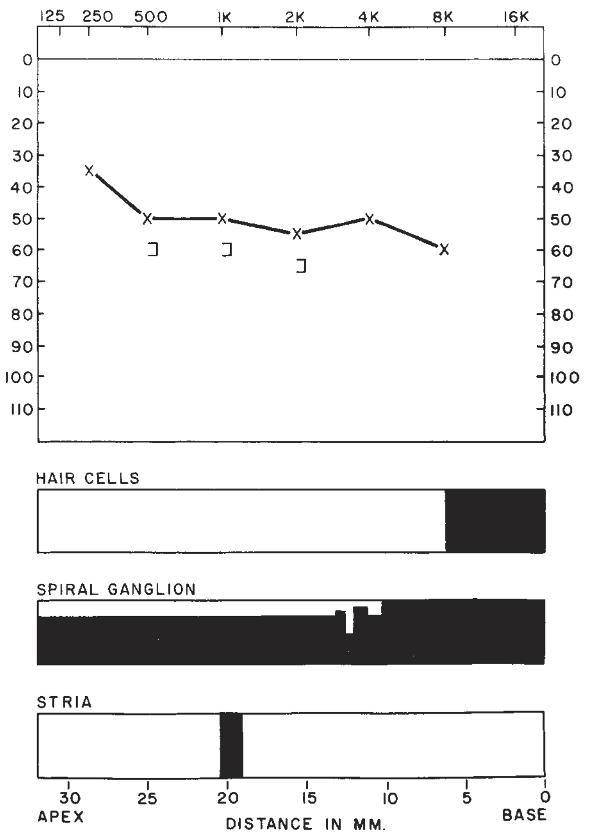
B

Figure 23–29 *Sensory presbycusis, right ear. This 83-year-old man had bilateral symmetrical progressive sensorineural loss, worse in the high frequencies.* ***(A)*** *There was loss of hair cells in the basal turn and presumed secondary cochlear neuronal degeneration.* ***(B)*** *There was also atrophy of the stria vascularis (magnification X56).*

Approximately one in three individuals over the age of 70 in the United States has a sensorineural loss. Undoubtedly, this represents a cumulative effect of several disorders that include noise trauma, ototoxicty, and genetically determined progressive degeneration. Presbycusis assumes four phenotypic patterns based on characteristic audiometric patterns and otopathology: atrophy of hair cells, neurons, and stria vascularis and an indeterminate group.

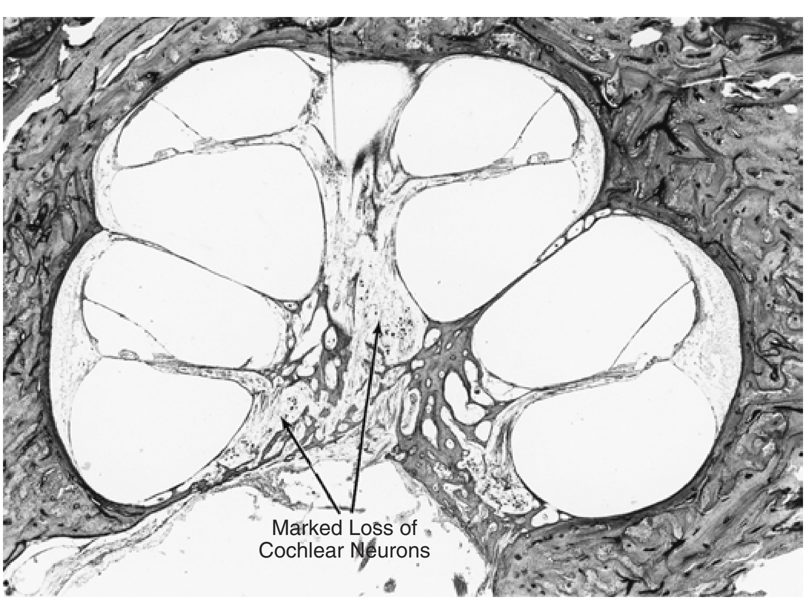
Sensory Presbycusis (Fig. 23–29)

This disorder is characterized by a sharply down-sloping bilateral symmetrical sensorineural hearing loss with



A

Figure 23–30 (A–C) *Neural presbycusis, left ear. This 63-year-old woman first developed a progressive bilateral sensorineural loss at age 38. There were normal populations of hair cells and supporting cells and a normal stria vascularis. However, there was marked degeneration of 85% of the cochlear neurons (magnification of* ***B*** *X27;* ***C*** *X66).*



B

C

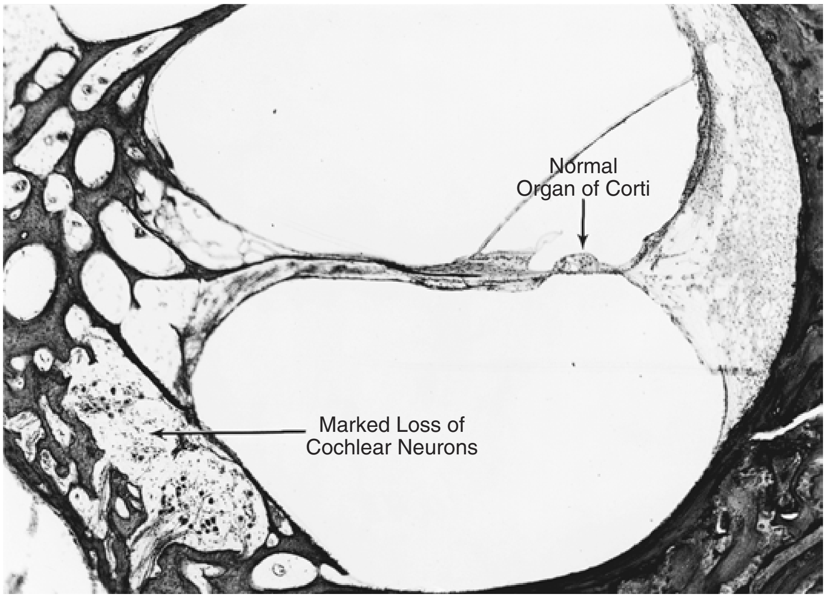


Figure 23–30 *(Continued)*

little effect on speech discrimination until the hearing loss becomes severe. The characteristic histopathology is degeneration of inner and outer hair cells, particularly in the basal turn.

Neural Presbycusis (Fig. 23–30)

The characteristic audiometric pattern of this disorder is a slowly descending audiometric pattern, worse in the high frequencies, and a significant decrement in speech discrimination. The characteristic otopathology is degeneration of primary cochlear neurons, particularly in the basal turn.

Strial Atrophy (Fig. 23–31)

The characteristic audiometric pattern is a flat sensorineural loss affecting all frequencies relatively equally, generally with maintenance of near normal speech discrimination. The characteristic otopathology is atrophy of the stria vascularis, which may vary among cochlear turns.

Indeterminate Presbycusis (Fig. 23–32)

In a small percentage of sensorineural hearing loss associated with the aging process, no convincing light microscopic otopathologic correlate can be found in the inner ear. Possible correlates of this hearing loss include CNS changes, submicroscopic anatomical or physiological disturbances in the inner ear, or dysfunction of supporting structures of the inner ear such as the spiral ligament or basilar membrane.

NEOPLASIA

The temporal bone may be the site of metastatic disease, the most common of which include primaries of the breast, kidney, and prostate, or may be the site of a primary neoplasia such as squamous cell carcinoma, glomus tumors, acoustic neuromas, and primarily tumors of the endolymphatic sac.

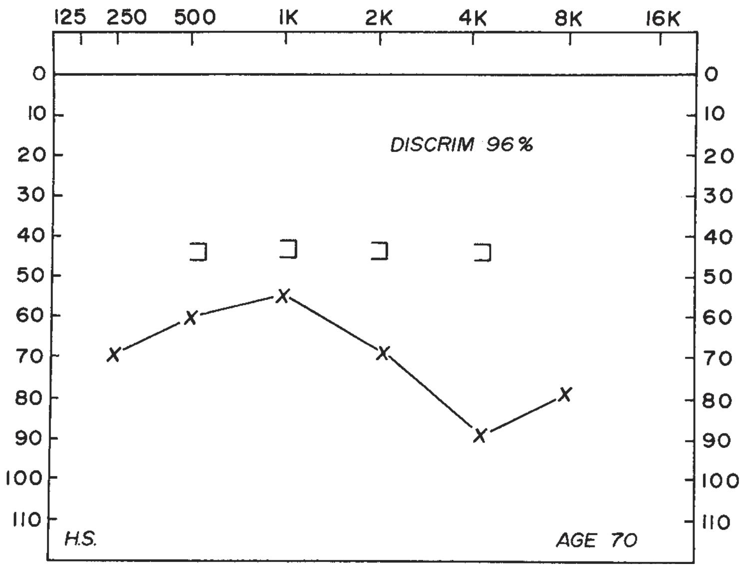
Squamous Cell Carcinoma (Fig. 23–33)

Squamous cell carcinoma of the auricle is clearly causally related to actinic exposure. Squamous cell carcinomas of the external auditory canal, middle ear, and mastoid are associated with a previous history of recurrent or chronic otorrhea, including external otitis and chronic otitis media. The cure rate of squamous cell carcinoma is markedly poorer with involvement of the middle ear and mastoid.

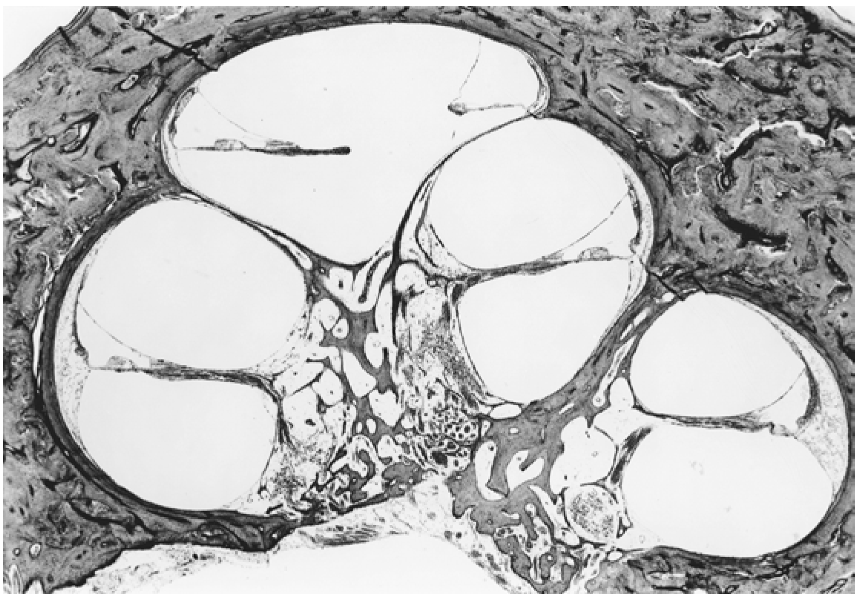
Acoustic Neuroma (Vestibular Schwannoma) (Fig. 23–34)

Vestibular schwannoma is one of the most common intracranial tumors. It occurs in a sporadic form or associated with neurofibromatosis type 2 (NF2). In NF2 there is an autosomal dominant pattern of inheritance caused by mutations on chromosome 22 resulting in failure of a suppressor gene.

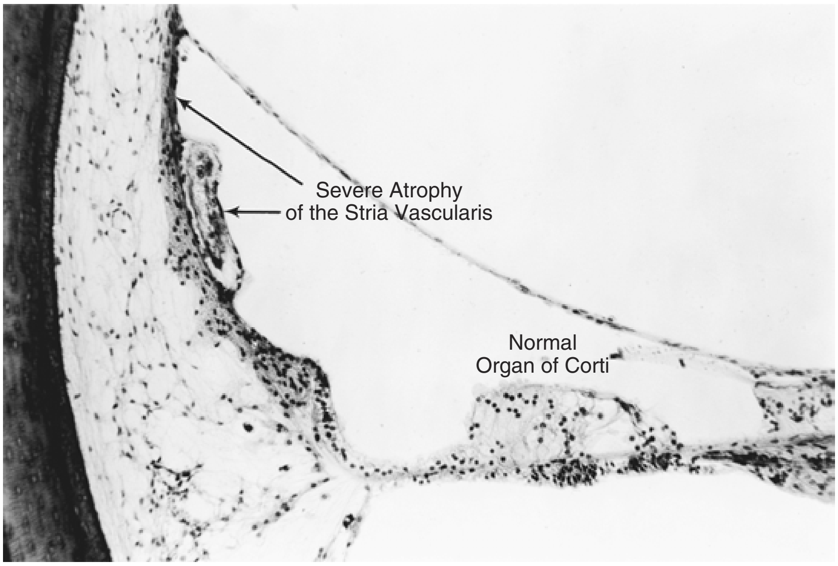
Vestibular schwannomas generally present with slowly progressive sensorineural hearing loss with marked decrement in speech discrimination. However, fluctuating or sudden sensorineural hearing loss



A



B



C

Figure 23–31 ***(A–C)*** *Presbycusis due to strial atrophy, left ear. This 72-year-old man first developed a bilateral progressive sensorineural loss in his 30s. An audiogram at age 70 demonstrated a mixed hearing loss, primarily sensorineural, with excellent speech discrimination. The conductive hearing loss was explained by otosclerotic fixation of the stapes bone. The primary histopathology of the inner ear was severe atrophy of the stria vascularis in all turns (magnification of* ***B*** *X27;* ***C*** *X150).*

A

B

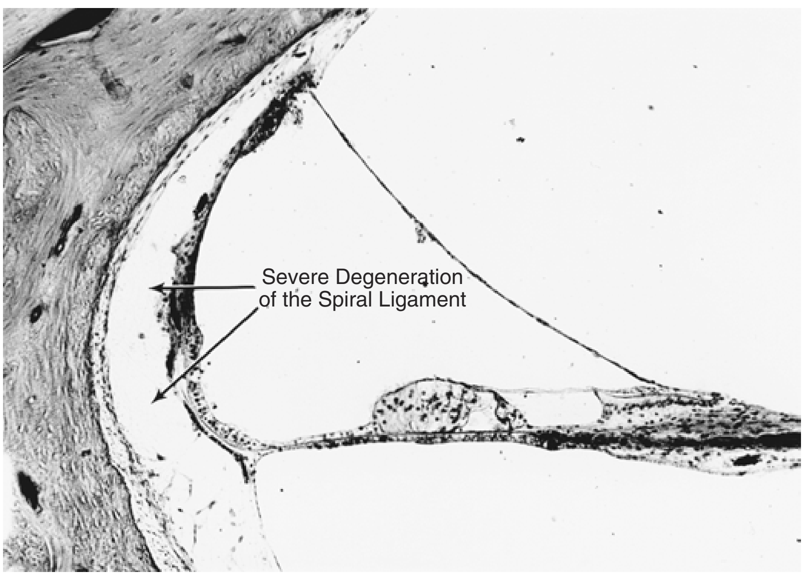


Figure 23–32 ***(A,B)*** *Possible inner ear conductive hearing loss secondary to degeneration of the spiral ligament, left ear. This 62-year-old man developed a progressive hearing loss starting at age 49. The organ of Corti was normal. There was partial loss of spiral ganglion cells in the basal turn and moderate degeneration of the stria vascularis. The principal histopathologic change was degeneration of the spiral ligament in the middle and apical turns and moderate degeneration of the spiral ligament in the basal turn (magnification of* ***A X****25;* ***B X****135).*

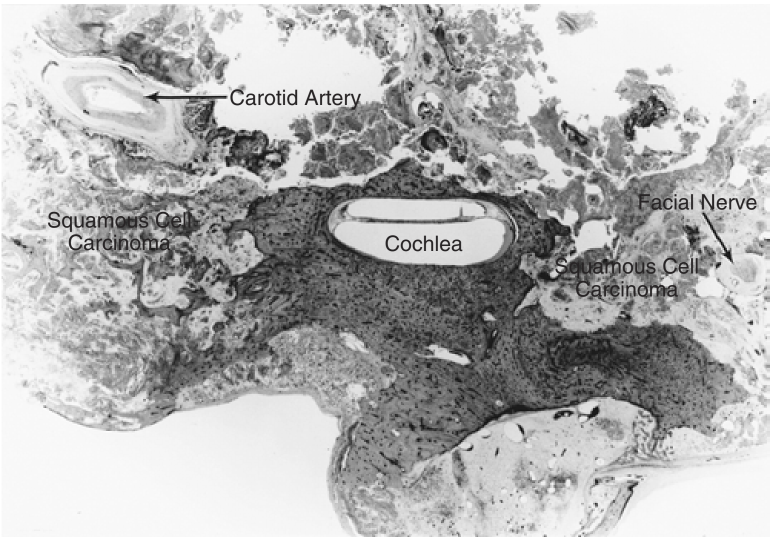
may occur. Vestibular complaints generally are mild and are seen in less than 50% of patients. Delayed complications of undetected vestibular schwannomas are due to cerebellar and brainstem compression and obstruction of CSF outflow.

Glomus Tumors (Fig. 23–35)

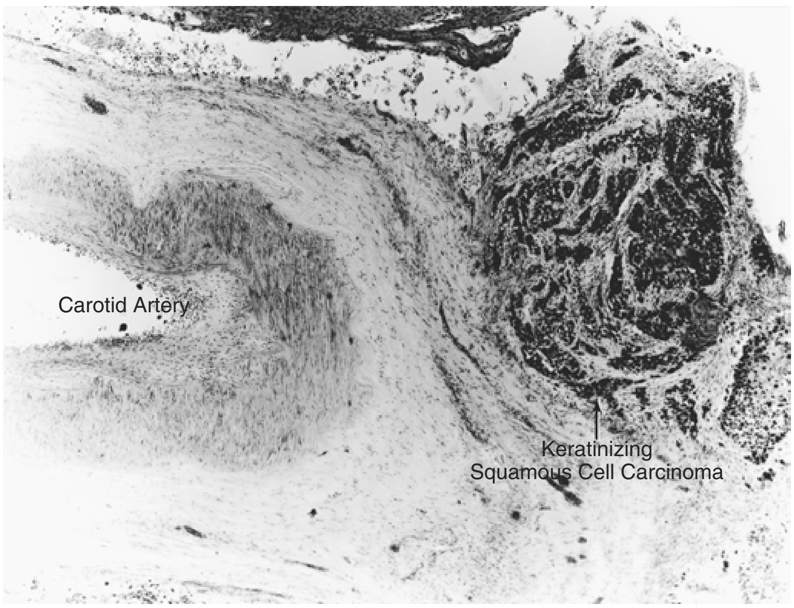
A glomus tumor (chemodectoma) may arise in the ear, head, and neck and mediastinal areas from normally occurring chemoreceptor organs. In the temporal bone, the most common sites of origin are the jugular bulb (glomus jugulare), promontory of the middle ear (glomus tympanicum), or vagus nerve (glomus juxtavagale). These tumors are almost always benign, but malignant transformation has been described. The most common clinical presentation includes pulsatile tinnitus, a vascular mass behind the tympanic membrane, and progressive dysfunction of cranial nerves VII and VIII. Intracranial and other cranial nerve involvement may be seen in larger tumors.

Endolymphatic Sac Tumors (Fig. 23–36)

Von Hippel-Lindau disease is a rare genetic disorder that has been linked to chromosome 3. A defect in a tumor suppressor gene is inherited as an autosomal dominant disorder and may result in pheochromocytomas, angiomas



A



B

Figure 23–33 (A,B) *Squamous cell carcinoma of the right temporal bone. This 70-year-old man, with a history of bilateral chronic otitis media since childhood, developed a squamous cell carcinoma in the right temporal bone at age 70. Symptoms included facial paralysis and a profound sensorineural loss on the right. Otologic examination revealed a large mass filling the right external auditory canal. Imaging demonstrated a lytic lesion in the inferior aspect of the right petrous bone, including the external auditory canal and jugular foramen. The patient died during the course of radiation therapy. A massive tumor destroying the temporal bone was visible on histopathology. The tumor surrounds the descending segment of the facial nerve and carotid artery (magnification of A X6; B X32).*

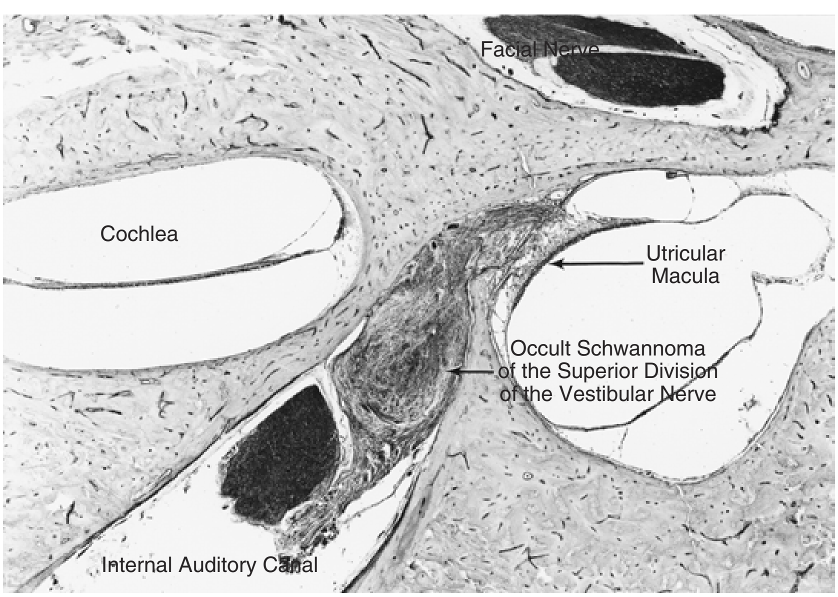
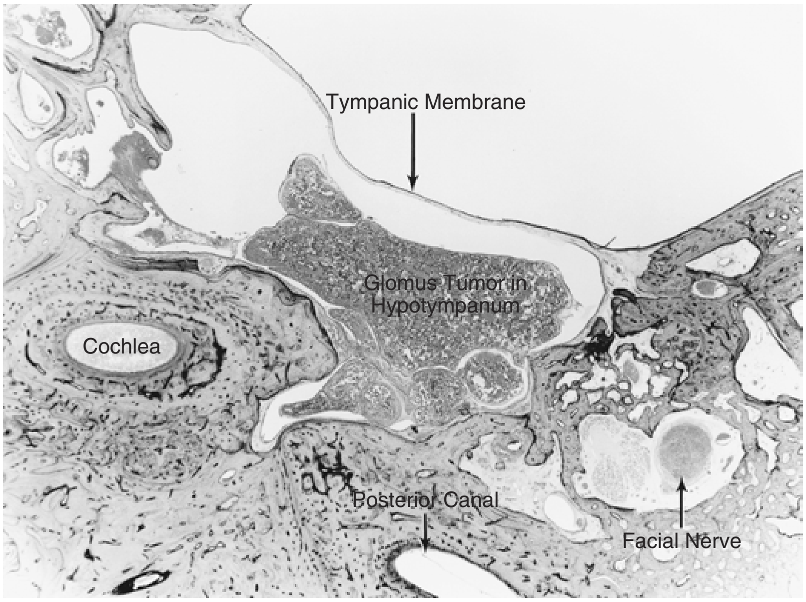
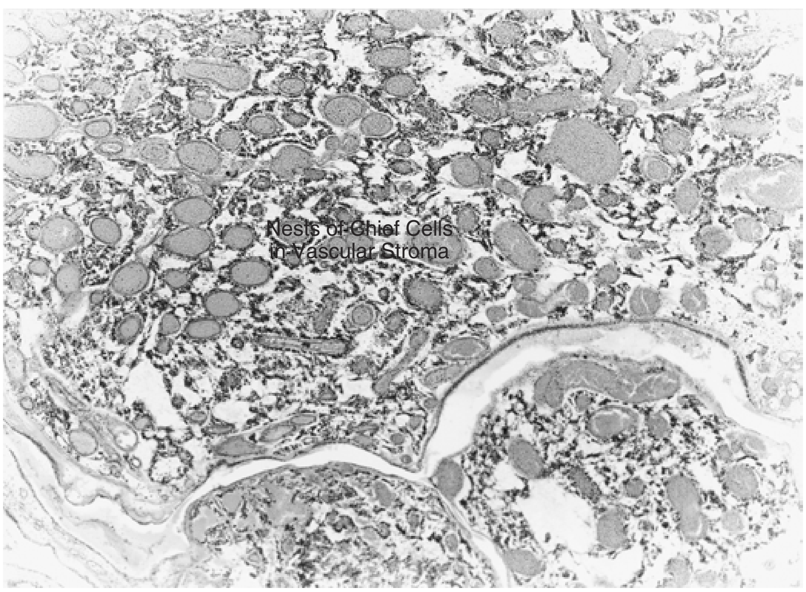


Figure 23–34 *Acoustic neuroma. This 84-year-old man had a bilateral mixed hearing loss with bilateral reduced speech discrimination, worse on the right side. The conductive component of his hearing loss was attributed to otosclerosis. Histologic study of the right temporal bone revealed an occult schwannoma of the superior division of the vestibular nerve in the distal internal auditory canal (magnification X18).*



A



B

Figure 23–35 ***(A,B)*** *Glomus tumor. This 89-year-old man was diagnosed with a left-sided posterior fossa meningioma at age 85. He had no symptoms related to his right ear and died at age 89 of renal failure. Histologic section of his right temporal bone revealed an occult glomus jugulare tumor extending into the hypotympanum (magnification of* ***A X****9.5;* ***B X****54.2).*

of the liver, kidney, and pancreas, hemangioblastomas of the CNS, and papillary cystadenomas of the endolymphatic sac. Occasionally, symptoms related to involvement of the endolymphatic sac may be the presenting symptoms and may mimic Meniere's disease.

Idiopathic Sensorineural Loss (Fig. 23–37)

Sudden idiopathic sensorineural hearing loss may occur with no systemic signs of other illness. It is generally unilateral and involves all age groups. The degree of hearing loss may be mild to profound and may involve concomitant vestibular symptoms. Although the etiology is unknown, the pathogenesis may include viral or immune etiologies. Some degree of recovery of mild to moderate sensorineural loss is common and is potentiated by the administration of corticosteroids.

Meniere's Syndrome (Fig. 23–38)

The classic triad of Meniere's syndrome includes fluctuant sensorineural hearing loss, ipsilateral aural fullness and tinnitus, and attacks of vertigo. Based on its distinctive otopathology, Meniere's syndrome is clearly a disorder of the inner ear. However, its etiology remains unknown. It is probably best understood as a syndrome rather than a disease; that is, a

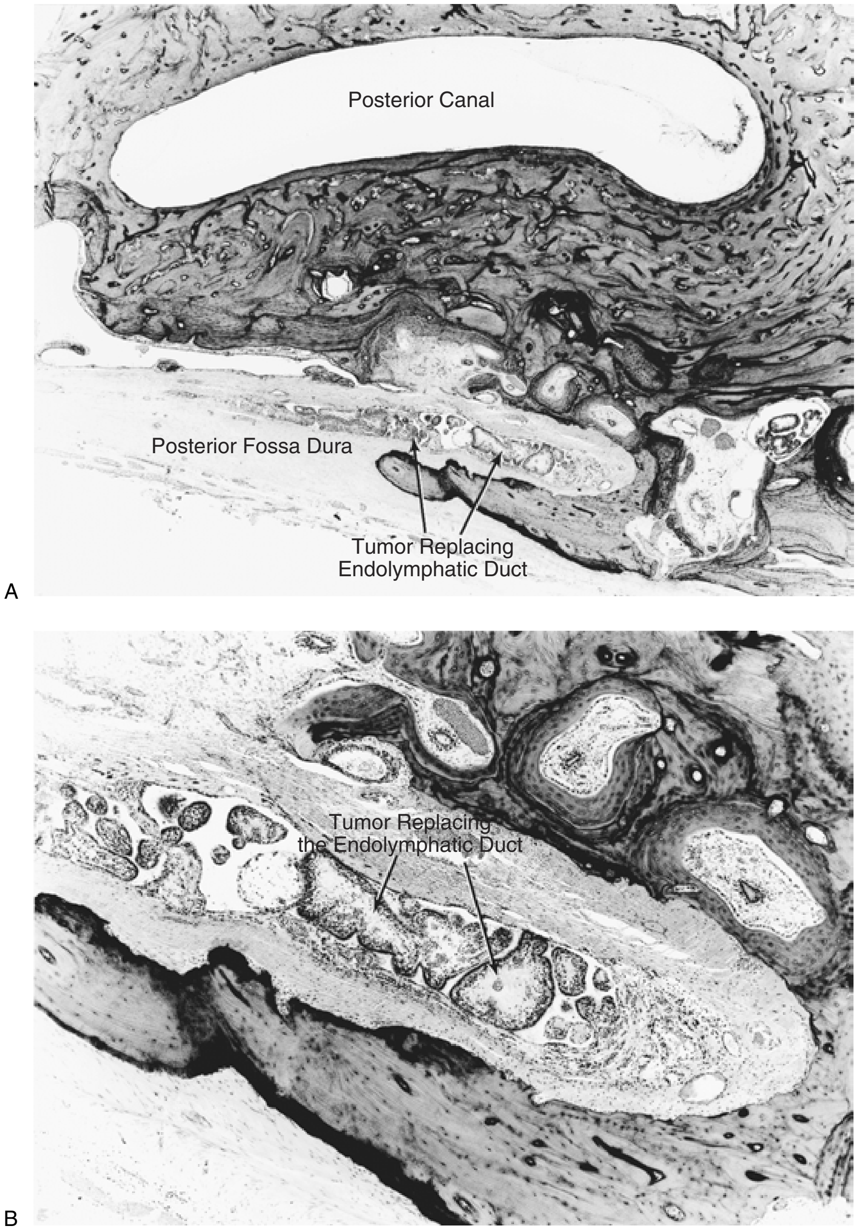


Figure 23–36 ***(A,B)*** *Endolymphatic sac tumor (papillary cystadenoma) in left ear. This 40-year-old patient with von Hippel-Lindau disease developed bilateral fluctuating and progressive sensorineural loss without vertigo at age 28. He had undergone bilateral endolymphatic sac decompressions based on a clinical diagnosis of Meniere's disease. He also had undergone exploratory tympanotomy for suspected perilymph fistulas. He died at age 40 of renal failure. Histopathology of the temporal bones demonstrated a bilateral papillary cystadenoma of the endolymphatic sac (magnification of* ***A*** *X20;* ***B*** *X58.8).*

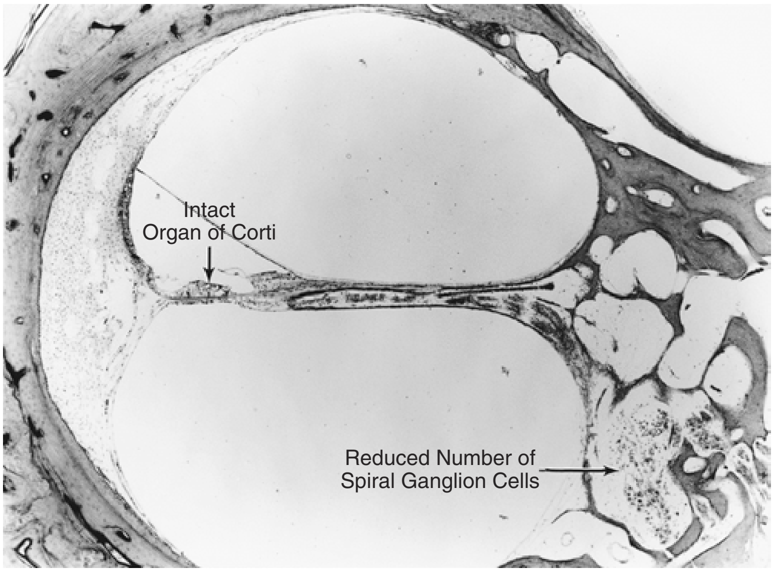


Figure 23–37 *Sudden idiopathic sensorineural hearing loss. This 67-year-old man developed a sudden subtotal sensorineural hearing loss of his right ear with vertigo at age 60. The pure-tone average was ~60 dB, and the speech discrimination was 12%. Imaging failed to demonstrate evidence of a retrocochlear lesion. Histopathologic examination of the right cochlea demonstrated reduction of the number of spiral ganglion cells in Rosenthal's canal and a normal organ of Corti. This was consistent with an idiopathic sudden sensorineural loss secondary to neuronal degeneration (magnification X53.8).*

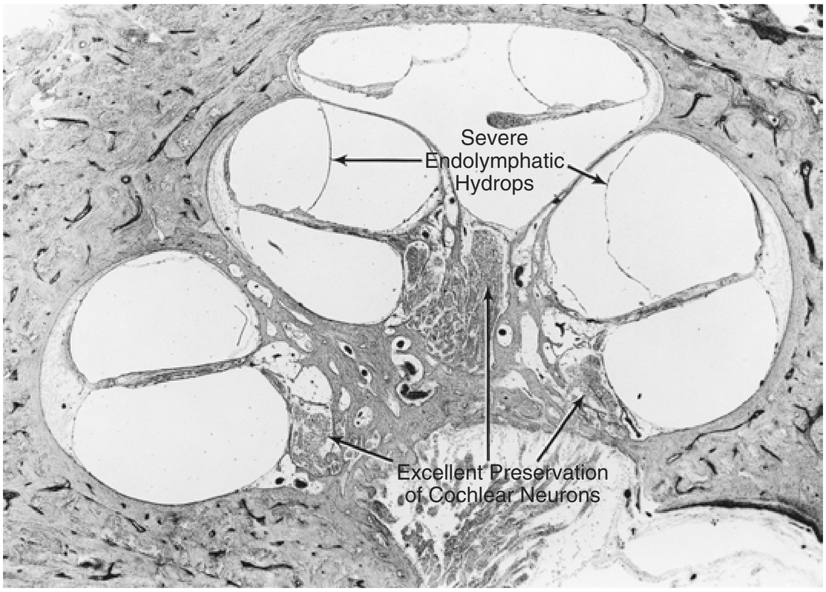


Figure 23–38 *Meniere's syndrome. This 53-year-old woman developed progressive sensorineural loss in her right ear and unsteadiness. Audiometry showed a severe sensorineural loss with 8% speech discrimination. Histopathology revealed severe endolymphatic hydrops of the right ear. As is common in Meniere's disease, there was no histopathologic correlate at a light microscopic level with the sensorineural loss on the right side (magnification X25).*

phenotypic clinical expression with the possibility of several possible etiologies. The classic otopathologic changes include the presence of endolymphatic hydrops and degeneration of hair cells and neurons, particularly in the apical turn of the cochlea.

Benign Paroxysmal Positional Vertigo (Fig. 23–39)

Benign paroxysmal positional vertigo (BPPV) is characterized by the induction of vertigo and rotatory nystagmus in the left or right head-hanging position on Hallpike testing. The nystagmus demonstrates latency and fatigue, which helps to differentiate it from more central forms of positional vertigo. It is due to pathological separation of otoliths from the utricular macula that find their way into the posterior or other semicircular canals. Once in the posterior canal, these free otoliths come in contact with and displace the cupula and mimic an angular acceleration of the head. These particles may lie free within the semicircular canal and be amenable to "repositioning maneuvers" to return them to utricular space or may become fixed on the cupula of the posterior semicircular canal.

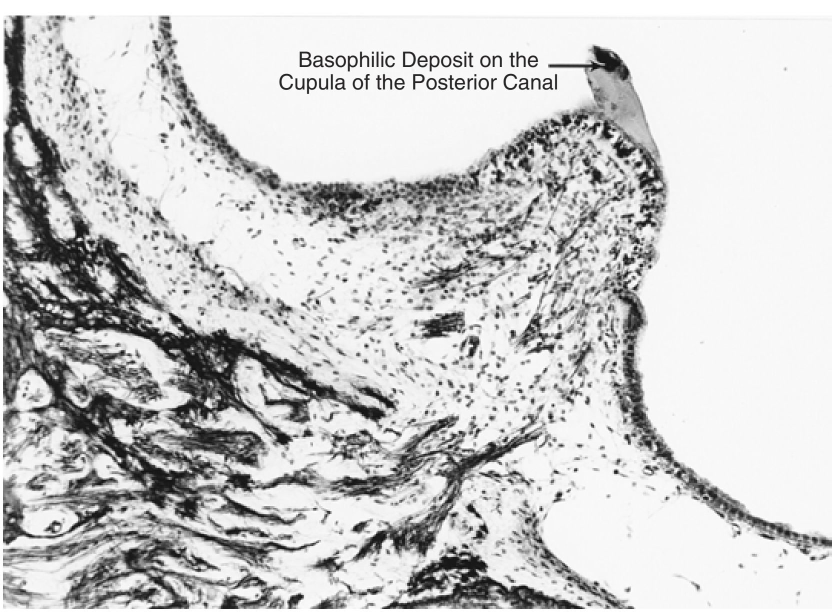


Figure 23–39 *Benign paroxysmal positional vertigo (BPPV). This 59-year-old man first developed symptoms consistent with BPPV at age 59. Histopathologic study of the posterior semicircular canal on the left side demonstrated a basophilic deposit fixed to the cupula (magnification X175).*

Suggested Readings

Gorlin RJ, Toriello HV, Cohen MM Jr. Hereditary Hearing Loss and Its Syndromes (Oxford Monographs on Medical Genetics No. 28). Oxford: Oxford University Press; 1995

Gorlin RJ, Webster DC, Carey JC. Genetic hearing loss associated with renal disorders. In: Gorlin RJ, Toriello HV, Cohen MM Jr, eds. Hereditary Hearing Loss and Its Syndromes (Oxford Monographs on Medical Genetics No. 28). Oxford: Oxford University Press; 1995:234–238

McKenna MJ, Kristiansen AG, Haines J. Polymerase chain reaction amplification of a measles virus sequence from human temporal bone sections with active otosclerosis. Am J Otol 1996;17(6):827–830

Merchant SN, Schuknecht HF, Rauch SD, et al. The national temporal bone, hearing and balance pathology resource registry. Arch Otolaryngol Head Neck Surg 1993;119: 846–853

Moscicki RA, San Martin JE, Quintero CH, Rauch SD, Nadol JB Jr, Bloch KJ. Serum antibody to inner ear problems in patient with progressive hearing loss. JAMA 1994;272:611–616

Nadol JB Jr. Application of electron microscopy to human otopathology. Acta Otolaryngol (Stockh) 1988;105:411–419

Nadol JB Jr. Techniques for human temporal bone removal: Information for the scientific community. Otolaryngol Head Neck Surg 1996;115:298–305

Schuknecht HF. Pathology of the Ear. 2nd ed. Philadelphia: Lea & Febiger; 1993

Wackym PA, Simpson TA, Gantz BJ, Suita RJ. Polymerase chain reaction amplification of DNA from archival celloidin embedded human temporal bone sections. Laryngoscope 1993;103(6):583–588

Wang PC, Merchant SN, McKenna MJ, Glynn RJ, Nadol JB Jr. Does otosclerosis only occur in the temporal bone? Am J Otol 1999;20(2):162–165

ACKNOWLEDGMENT

My sincere appreciation to Barbara Burgess and Jessica Hutta for their help and painstaking attention to detail in the creation of the illustrations.

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. For the human temporal bone, the most common plane of section is
   A. Coronal
   B. Sagittal
   C. Axial
   D. Waters
2. Von Hippel-Lindau disease may cause otologic symptoms mimicking
   A. Meniere's disease
   B. Benign positional vertigo
   C. Presbycusis
   D. Sudden idiopathic deafness
3. In the human, the most commonly affected cells of the inner ear in noise trauma are
   A. Intermediate cells of the stria vascularis
   B. Inner hair cells
   C. Interdental cells
   D. Outer hair cells

Chapter 24

ULTRASTRUCTURAL ANATOMY OF THE COCHLEA

DAVID J. LIM

The primary function of the cochlea is hearing. The cochlea (meaning "snail" in Latin) is composed of a sensory organ (organ of Corti), structural parts critical for the mechanical transmission and amplification of sound, and cells that participate in inner ear homeostasis (**Fig. 24–1**). The organ of Corti consists of both sensory and supporting cells. The organ of Corti, together with two membranous structures, the tectorial membrane and the basilar membrane, make up a sensory transduction complex needed for peripheral auditory function, as described in Chapters 25 and 26.

There are 26 cell types in the cochlea, but their specific functions (except in only a few cases) are not well established. We are now beginning to understand the roles that these cells play in both sensory transduction and inner ear homeostasis, on a cellular and molecular level. Of the 26 cell types, the majority of the cells are involved in supporting (both mechanically and functionally) the sensory receptor cells (**Fig. 24–2**). This functional support includes maintenance of the inner ear ionic and fluid homeostasis. The stria vascularis is generally accepted as being responsible for maintaining the high potassium (K^+) ion

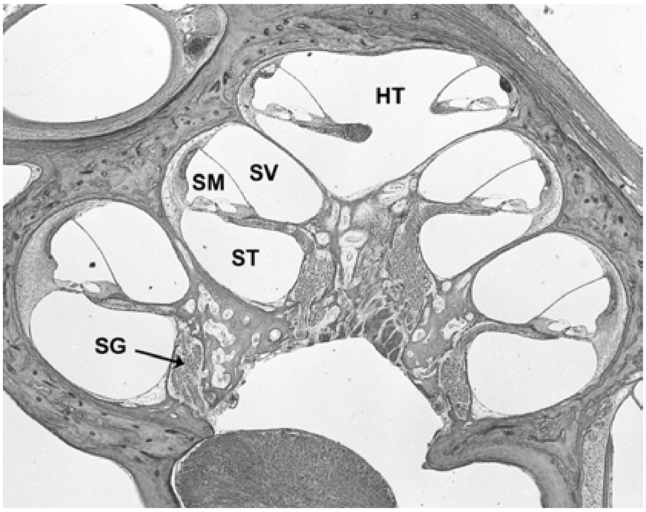


Figure 24–1 *A midmodiolar section of a squirrel monkey cochlea showing three turns with three distinct compartments: scala media (SM), filled with endolymph, and scalae vestibuli (SV) and tympani (ST), filled with perilymph. Clusters of spiral ganglions (SG) are found inside the Rosenthal's canal. Scala vestibuli and scala tympani open to each other at the apex of the cochlea known as the helicotrema (HT). Observe that the size of the organ of Corti is bigger at the apex and smaller at the base, like the basilar membrane and tectorial membrane, which are wider at the apex and narrower at the base. The size of the lateral wall including stria vascularis and spiral ligament is bigger at the base and smaller toward the apex. (Courtesy of D.J. Lim.)*

concentration of the endolymph and generating high endocochlear potential. It is now believed that the stria vascularis is an integral part of the complex pathways of ion homeostasis of the cochlea. Modern cell and molecular biology and gene manipulation have provided new insights into the variety of cellular functions among the cochlear nonsensory cells involved in ion transport, K^+ recycling, pH regulation, water transport, and innate immune activation. These functions are critical for sensory cell function as well as maintenance of the health of the sensory organ.

ORGAN OF CORTI

The uniqueness of the organ of Corti's anatomy is the tonotopic structural arrangement to provide a first filter (tuned) for detection of sounds of different frequencies. Fine-tuning capability is now known as the active hearing process and is provided exclusively by the outer hair cells themselves (see Chapters 25 and 26). The size of the organ of Corti varies gradually from base to apex: the sensory organ at the base is smaller, while the organ in the apical region is larger. The basilar and tectorial membranes also vary in width and thickness along the length of the cochlea: they are thicker and wider toward the apex and thinner and narrower in the base (**Fig. 24–1**). The tectorial membrane is a gelatinous membrane whose undersurface firmly anchors the tips of the tall rows of outer hair cells. This anatomical coupling is crucial for deflection of the stereociliary bundles and is required for mechanosensory transduction. (See Tectorial Membrane section.)

Sensory Cells

Corti's organ consists of a single row of inner hair cells (IHCs) and three rows of outer hair cells (OHCs) in mammals (**Figs. 24–3, 24–4**). There are ~4,000 IHCs

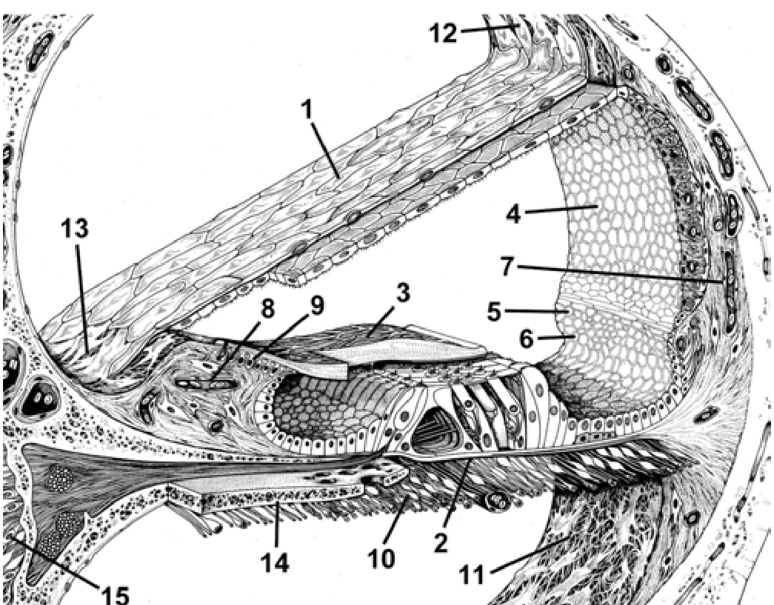


Figure 24–2 *An artistic rendering of a section of cochlea in the basal turn showing (1) Reissner's membrane, (2) basilar membrane, (3) tectorial membrane, (4) stria vascularis, (5) spiral prominence, (6) external sulcus cells, (7) spiral ligament, (8) spiral prominence, (9) interdental cells, (10) mesothelial cell of basilar membrane, (11) infrabasilar membrane area of spiral ligament, (12) suprastrial area of spiral ligament, (13) supraspiral prominence area of spiral limbus, (14) osseous spiral lamina, and (15) spiral ganglion. (Courtesy of D.J. Lim; drawing by Nancy Sally.)*

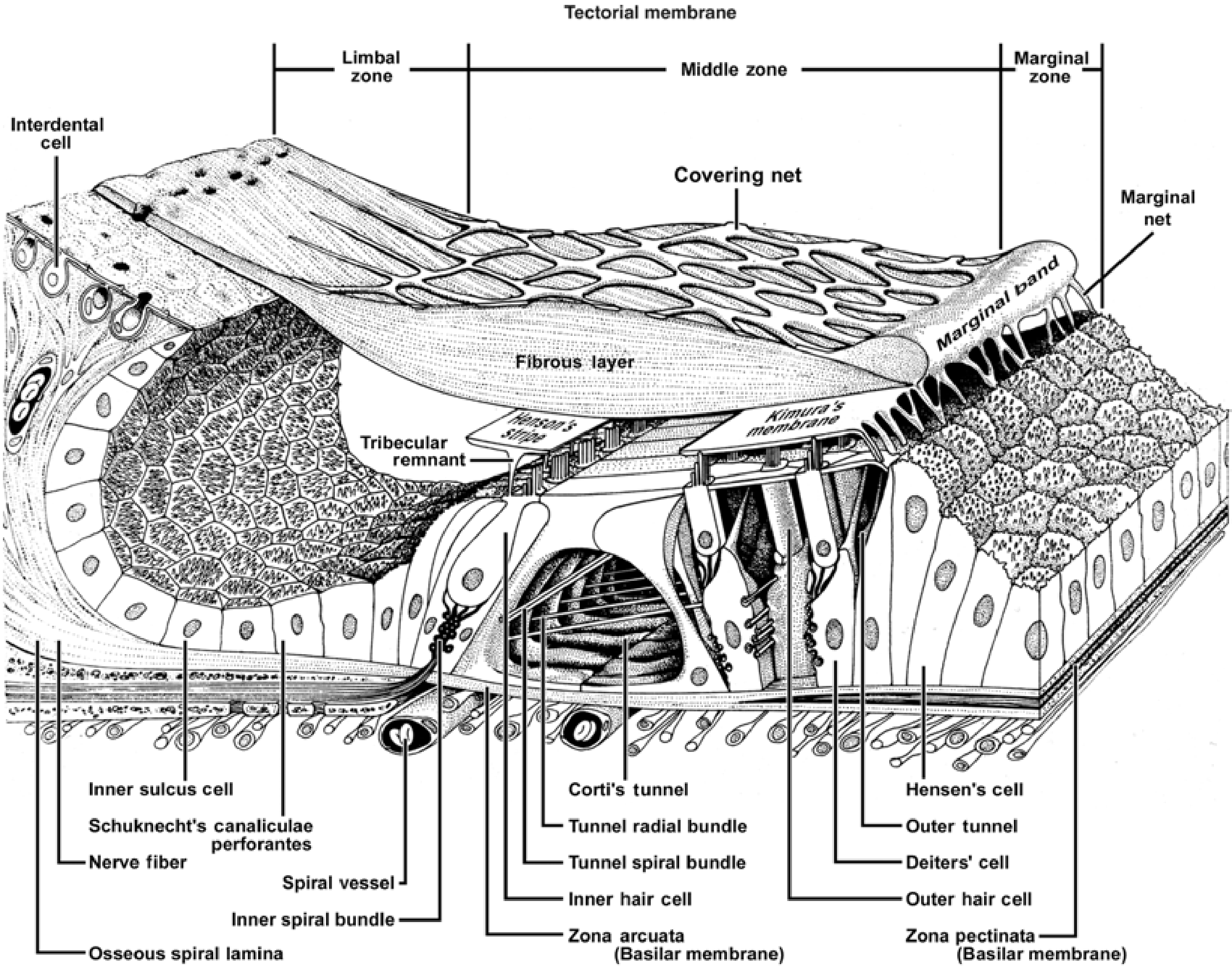


Figure 24–3 *An artist's concept of the organ of Corti showing details of the substructures of the tectorial membrane, basilar membrane, and organ of Corti. (From Lim D.J. Functional structure of the organ of Corti: a review. Hear Res 1986;22:117–146. Reprinted with permission. Drawing by Nancy Sally.).*

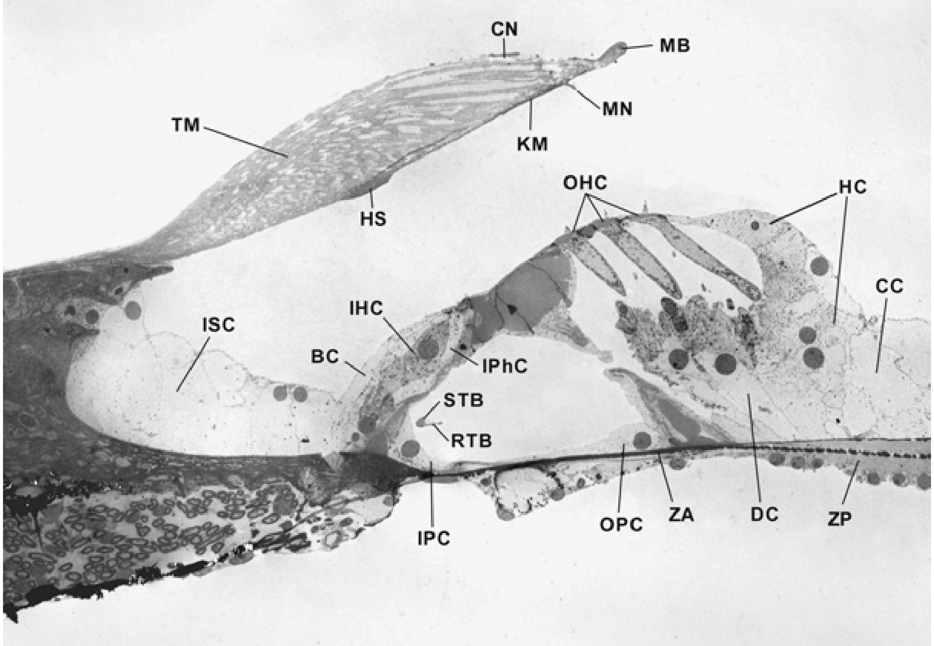


Figure 24–4 *A transmission electron microscopy (TEM) photomicrograph of chinchilla organ of Corti, tectorial membrane (TM), and basilar membrane showing the following: Hensen's stripe (HS), Kimura's membrane (KM), marginal net (MN), marginal band (MB), cover net (CN), inner sulcus cell (ISC), border cell (BC), inner hair cell (IHC), inner phalangeal cell (IphC), inner pillar cell (IPC), outer pillar cell (OPC), outer hair cell (OHC), Deiters' cell (DC), Hensen's cell (HC), Claudius' cell (CC), zona accuata (ZA) and zona pecta (ZP) of basilar membrane, and Corti's tunnel with spiral tunnel nerve bundle (STB) and radial tunnel nerve bundle (RTB). (Courtesy of D.J. Lim.)*

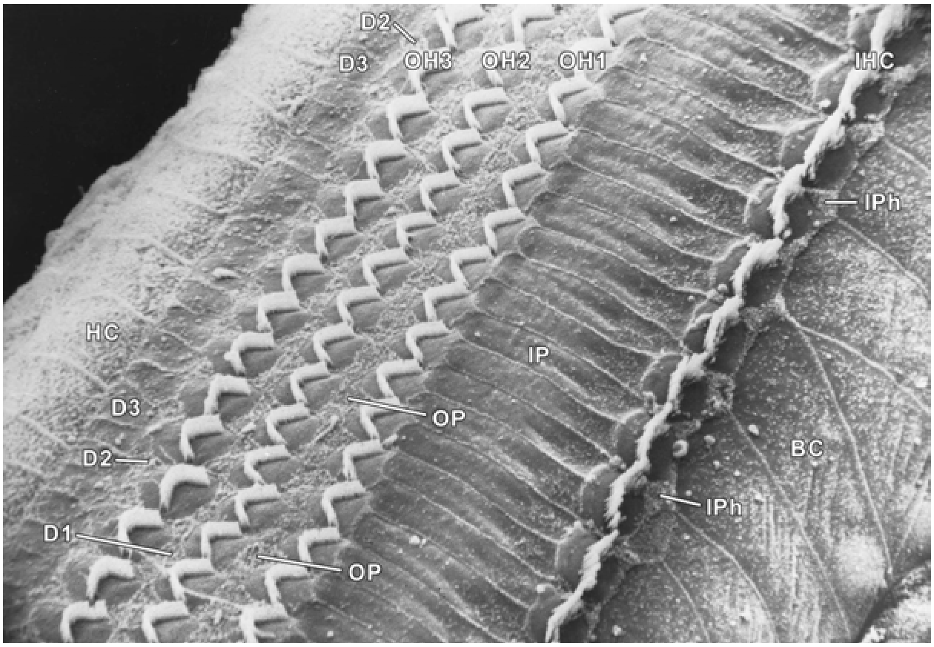


Figure 24–5 *Surface view of chinchilla organ of Corti showing inner phalangeal cell (Iph), inner hair cell (IHC), inner pillar cell (IP), outer pillar cell (OP), and three rows of outer hair cells (OH1, OH2, OH3) and Deiters' cells (D1, D2, D3). HC, Hensen's cell. (Adapted from Lim, D.J. Functional structure of the organ of Corti: a review. Hear Res 1986;22:117–146. Reprinted with permission.)*

and 12,000 OHCs in the human cochlea. The surface of the sensory cells bears stereociliary bundles facing the endolymph of the scala media. Each of the stereociliary bundles is formed by 60 to 120 cilia of varying length in a W-shaped, steplike arrangement (**Figs. 24–5, 24–6, 24–7**). This arrangement is critical for mechanotransduction because each ciliary tip is linked by a twisted ropelike, fine filamentous "tip link" to the neighboring taller stereocilia, establishing directional sensitivity of the sensory ciliary bundles (**Figs. 24–8, 24–9**). Stereocilia are made up of a paracrystalline array of actin filaments, espin, and myosin VI and VIIA, and these molecules undergo continuous turnover. Actin and espin are incorporated at the paracrystalline tip of the stereocilia and flow downward at a rate proportional to the length of stereocilia. In other words, the entire staircase (made up of three rows) of stereociliary bundles is turned over synchronously. Details of the transduction apparatus are covered in Chapter 25.

There are several genetic mutations that affect the formation and maintenance of the stereocilia (see Chapter 19). *MYO7A* gene, encoding myosin VIIA, shows

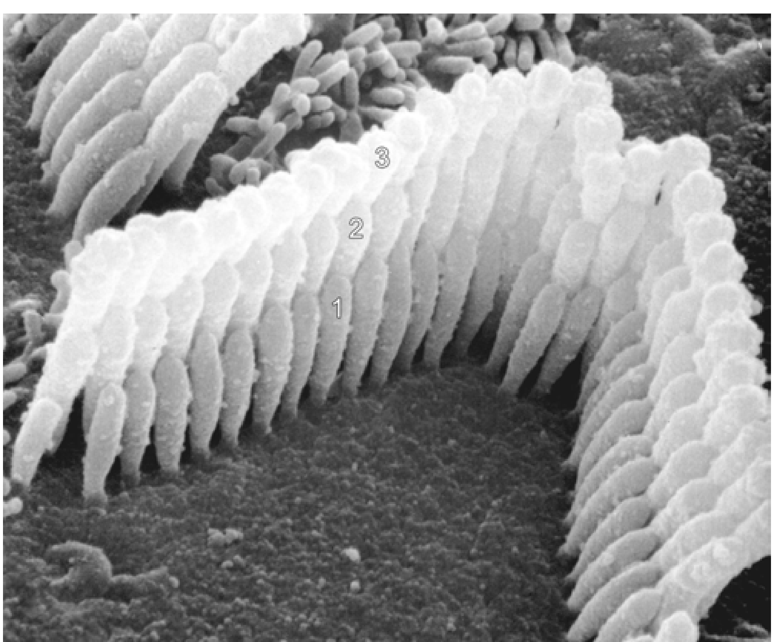


Figure 24–6 *Close-up view of stereociliary bundle of a first-row outer hair cell from the middle turn of a cochlea (chinchilla), showing the steplike arrangement of the stereocilia of different heights (first, second, third rows). Because of the indentation of the angulated portion where the kinocilium was present during development but later atrophies, the ciliary bundle forms a W pattern. Observe the rounded tips of the third row of the stereocilia and the pointed tips of the lower first and second rows of stereocilia. The pointed tips of these first- and second-row stereocilia are a result of where the tip links are attached. (From Lim DJ. Functional structure of the organ of Corti: a review. Hear Res 1986;22:117–146. Reprinted with permission.)*

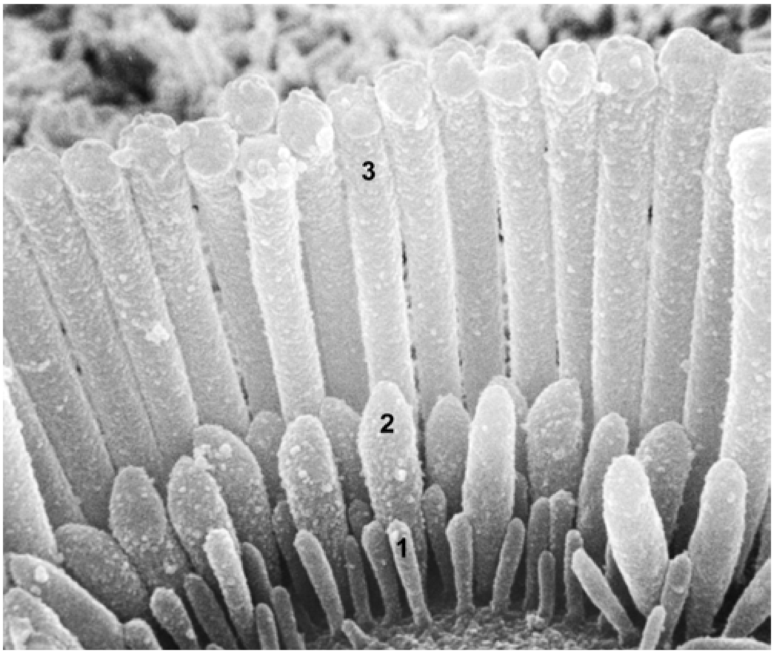


Figure 24–7 *Close-up view of stereociliary bundle of the inner hair cell of a cochlea (chinchilla) showing wide* W *formation with one tall row (third) and two shorter rows (first, second) of stereocilia with steplike arrangement. Note the round appearance of the stereociliary tips of the first row but the pointed tips of the second row of stereocilia. (From Lim, D.J. Functional structure of the organ of Corti: a review. Hear Res 1986;22:117–146. Reprinted with permission.)*

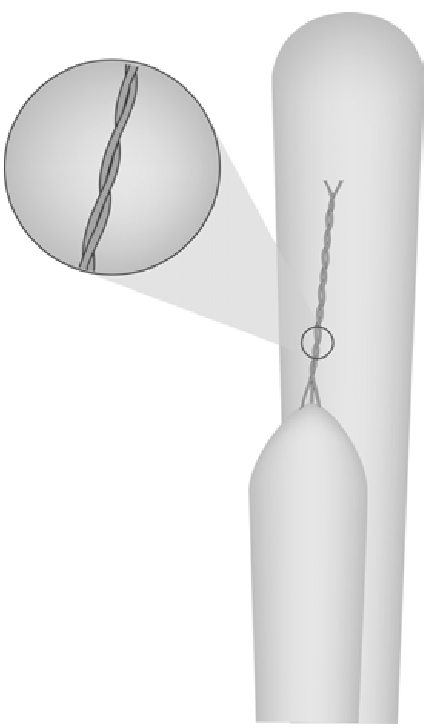


Figure 24–8 *An artistic conception of a twisted ropelike tip link connecting the tip of the lower row stereocilium to the side of a taller stereocilium. (Adapted from Kachar B, Parakkal M, Kurc M, Zhao Y, Gillespie, P.G. High-resolution structure of hair-cell tip links. Proc Natl Acad Sci USA 2000;97:13336–13341. Reprinted with permission. Drawing by Ken Sakai.)*

Figure 24–9 *Two tip links (arrows) of outer hair cells of a guinea pig cochlea, prepared by the freeze-fracture replica technique, showing a bifurcated attachment of the tip links from the tops of the shorter stereocilia to the sides of the taller stereocilia. (Adapted from Kachar B, Parakkal M, Kurc M, Zhao Y, Gillespie, P.G. High-resolution structure of hair-cell tip links. Proc Natl Acad Sci USA 2000;97:13336–13341. Reprinted with permission.)*

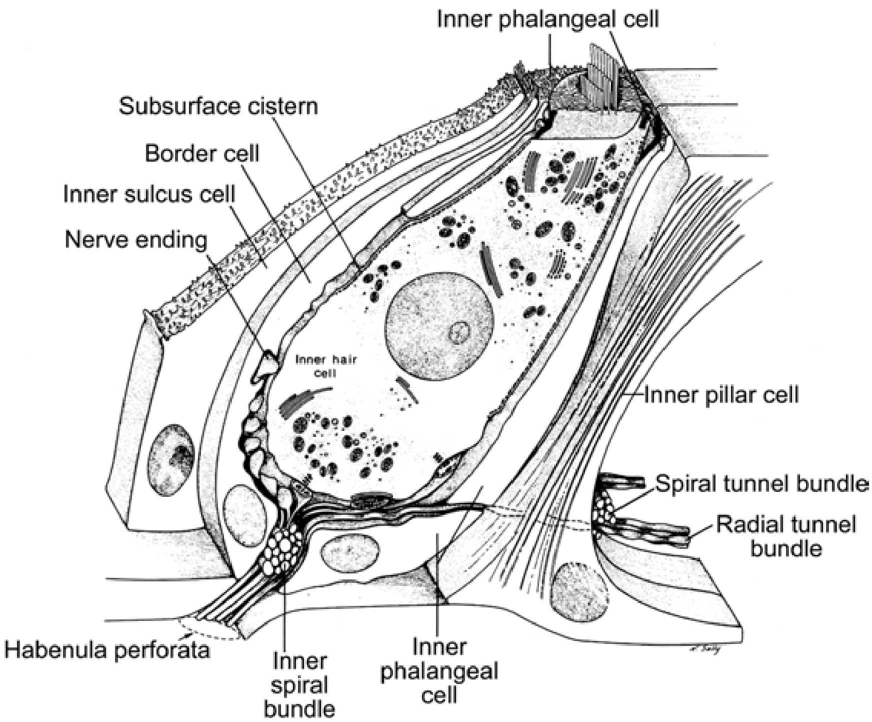


Figure 24–10 *An artist's conception of an inner hair cell, showing the organization of cell organelles that is distinct from that of an outer hair cell depicted in Fig. 24–11. (From Lim, D.J. Functional structure of the organ of Corti: a review. Hear Res 1986;22:117–146. Reprinted with permission.)*

defects in morphogenesis of the inner ear sensory cell stereocilia. *WHRN* gene, encoding calmodulin-dependent serine kinase (CASK)–interacting protein 98 or whirlin, is involved in the elongation of the stereocilia. The "Whirler" mouse is deaf and shows stereociliary abnormalities. It is suggested that whirlin is involved in the elongation and maintenance of the stereocilia in both the IHCs and the OHCs.

The ultrastructural organizations of the inner and outer hair cells have the common feature of their stereociliary complexes, although they have distinctly different cell bodies. The stereociliary complex includes well-organized steplike stereociliary bundles, membrane-bound nonspecific mechanotransduction channels attached to filamentous tip links, and side links of stereocilia and cuticular plates. The exact nature of these links is not yet characterized.

The cell body of the IHC contains numerous vesicles and mitochondria (**Fig. 24–10**). Efferent nerve endings that make axodendritic synapses are attached to the basal part of the cell body. It is generally believed that IHCs are the primary mechanoreceptors that transmit neural signals to the auditory neurons of the spiral ganglion and the auditory centers of the brain, whereas the OHCs are responsible for sharp tuning and amplification of auditory signals, a process known as active hearing.

The OHC can modulate its length in response to auditory frequencies and is responsible for the active hearing process and for otoacoustic emissions (see Chapters 28 and 30). It is now known that the OHCs possess both calcium- and energy-dependent slow and calcium-independent fast motilities. Recent findings suggest that the cylindrical cell shape and motor activity of the hair cell are associated with the lateral wall of the OHC and that fast motility is the function of the OHC cell membrane, particularly the membrane protein known as prestin (**Figs. 24–11, 24–12**). The cell body of the outer hair cell contains a well-organized endoplasmic reticulum (ER) leaflet known as subsurface cisternae (SSC), a specialized ER, particularly along the lateral wall of the cell membrane (**Fig. 24–11**). These membranes are closely attached together through a well-developed cytoskeletal cortex consisting of actin and spectrin (**Fig. 24–12**). Recently, freeze-fracture replica technique allowed a detailed three-dimensional ultrastructural understanding of the complex structures of the cytoskeletal cortex and specialized ER membranous networks attached to the lateral cell membrane (**Figs. 24–11, 24–12, 24–13**). Numerous mitochondria are found tightly attached to the surface of the innermost leaflet of the SSC. The SSC become blind end-forming subsynaptic cisternae at the nerve-ending pole of the OHC, and efferent nerve endings attach to the outer hair cell apposing this region, indicating efferent control of OHC function (**Fig. 24–11**).

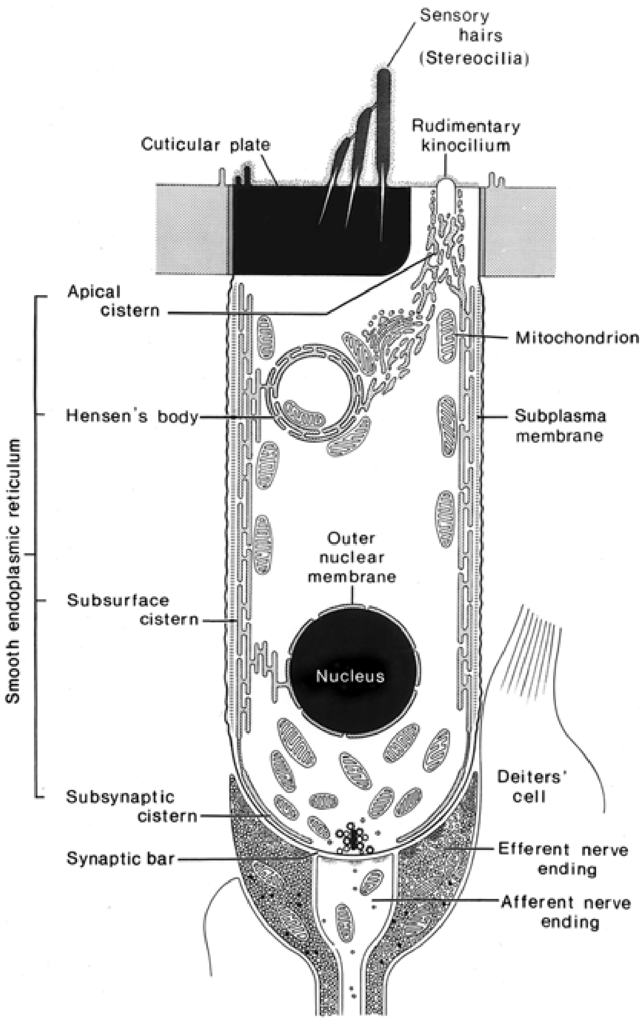


Figure 24–11 *A schematic diagram of an outer hair cell showing various cell organelles. Note the interconnecting cisternae structures, such as apical cistern, Hensen's body, subsurface cistern, and subsynaptic cisternae. (From Lim, D.J. Functional structure of the organ of Corti: a review. Hear Res 1986;22:117–146. Reprinted with permission. Drawing by Nancy Sally.)*

Supporting Cells (Pillar Cells, Deiters' Cells)

The supporting cells of the organ of Corti provide both structural and functional support. The inner pillar and outer pillar cells provide the structural pillars necessary to maintain the unique shapes of the organ of Corti and of Corti's tunnel (**Fig. 24–2**). The head plate of the pillar cells forms part of the reticular lamina together with phalanges of Deiters' cells. The rigidity of pillar cells is maintained by bundles of tubulin and intermediate filaments. Deiters' cells are made up of three components: a cylindrical body with an indented cup to hold the lower end of the outer hair cell, a slender cell process, and an umbrella-like phalanx that holds and seals the top portion of the OHCs and constitutes a major portion of the reticular lamina. Deiters' cells are extensively interconnected with the outer hair cell body and establish electrical coupling with other Deiters' cells. Inner hair cells, on the other hand, are surrounded by supporting cells that include inner phalangeal cells and inner border cells (**Fig. 24–2**).

It has been generally thought that these supporting cells provide mechanical support for the sensory cells and provide a mechanical link to the basilar membrane. However, the results of recent studies provide evidence that these cells also perform many important functions for the homeostasis of the inner ear and proposed K^+ recycling.

Hensen's Cells

Hensen's cells are tall columnar cells found next to the third row of Deiters' cells and are considered to be a part of the organ of Corti (**Fig. 24–2**). Hensen's cells can occur in several rows and sometimes give the appearance of a stratified formation, but it is not clear whether or not the cells under Hensen's cells are Claudius' cells that migrated during development, or are stratified Hensen's cells, distinct from Claudius' cells. In the guinea pig cochlea, this cell is characterized by large numbers of lipid droplets. In the cochlea of gerbils and bats, the first rows of Hensen's cells appear morphologically different from the rest of Hensen's cells, and thus are called "cover" or tectal cells. Because of the finger-like projections of their basolateral cell membranes it is suggested that this cell type is involved in regulation of the fluid and ion content of the Corti's outer tunnel. It is also believed that Hensen's cells are a part of the K^+ recycling pathway. There is some experimental evidence to indicate that Hensen's cells, like Deiters' cells, may be involved in the calcium buffer system. A recent study demonstrated that the water channel protein aquaporin-4 (AQP4) is expressed in the basolateral surface of Hensen's cells as well as in Claudius' cells and inner sulcus cells (**Fig. 24–14**). The AQP4 null mutant mouse is deaf (see Aquaporins section). It is postulated that AQP4 may be involved in the reabsorption of water and plays a role in maintaining the ionic balance of the inner ear fluids.

CELLS COVERING THE FLOOR OF THE SCALA MEDIA

Claudius' Cells

Claudius' cells are cuboidal cells covering the surface between Hensen's cells and external sulcus cells (root cells) (**Fig. 24–2**). These cells show strong expression of epithelial sodium Na^+ channel (ENaC), suggesting

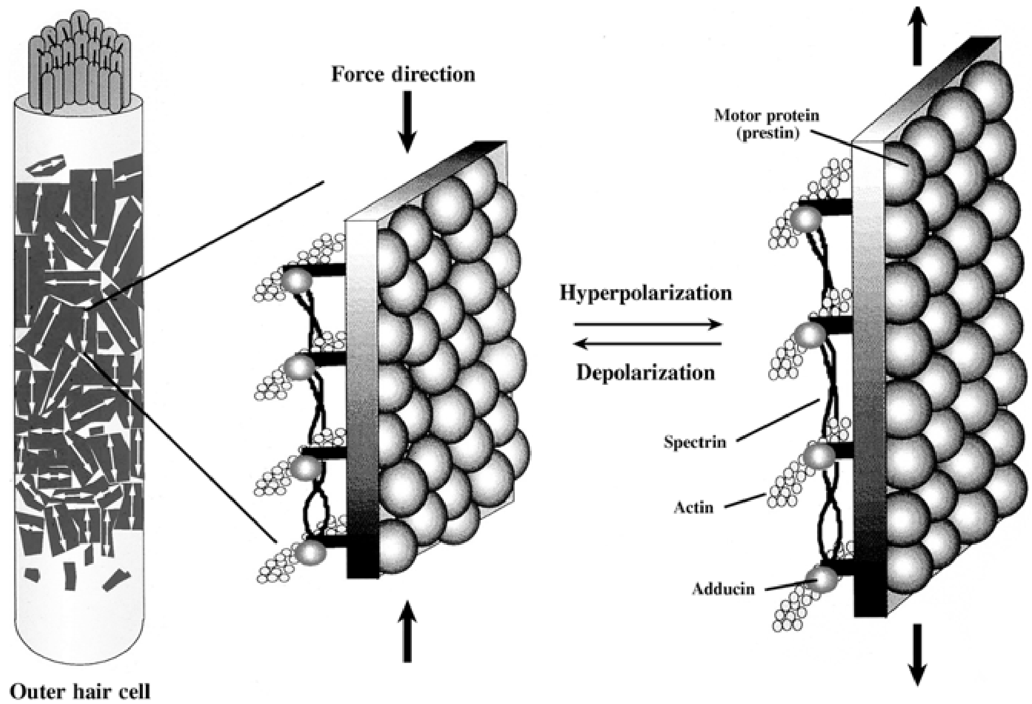


Figure 24–12 ***(A)*** *"Microdomains" model of the force generation mechanism acting during depolarization and hyperpolarization in the lateral plasma membrane of the outer hair cells. (From Lim, D.J, Kalinec F. Cell and molecular basis of hearing. Kidney Int Suppl 1998;65:S104–113, and adapted from Kalinec F, Kachar B. Structure of the electromechanical transduction mechanism in mammalian outer hair cells. In: Flock A, Ottoson D, Ulfendahl M, eds. Active Hearing. Amsterdam: Elsevier Science; 1995:179–191. Reprinted with permission.)*

their involvement in sodium ion reabsorption from the endolymph. Because of AQP4 expression in this cell type (**Fig. 24–14**), they are also most likely involved in fluid regulation of the inner ear. In addition, both Hensen's cells and Claudius' cells are also part of the K^+ ion-recycling pathway through gap junctions (see Gap Junctions).

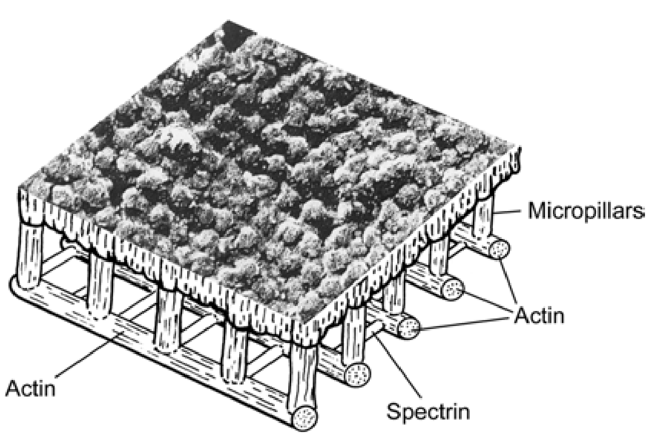


Figure 24–13 *An image of the transmembrane particles in the outer hair cells of the lateral plasma membrane is superimposed on a diagram of the cortical lattice. (Adapted from Kalinec F, Holley MC, Iwasa KH, Lim DJ, Kachar B. A membrane-based force generation mechanism in auditory sensory cells. Proc Natl Acad Sci USA 1992;89:8671–8675. Reprinted with permission.)*

Böttcher's Cells

Böttcher's cells are found sandwiched between the basilar membrane and Claudius' cells. They are found mostly in the basal turn and to a lesser extent in the middle turn, but they are not found in the apical turn. Böttcher's cells are cuboidal in shape and arranged in a single layer; they contain a rich complement of cell organelles such as endoplasmic reticulum, mitochondria, and microtubules. There is extensive microvilli formation along the lateral surface of these cells that form spiral intracellular channels. These cells also exhibit a high level of acid phosphatase, suggesting that they are secretory cells. They also contain carbonic anhydrase, and thus they may be involved in HCO_3 regulation.

Inner Sulcus Cells

The cuboidal inner sulcus cells resemble Claudius' cells with a relatively clear cytoplasm. These cells establish gap junctions with neighboring cells and are believed to be a part of the proposed medial K^+-recycling pathway to remove K^+ from the area of the IHCs.

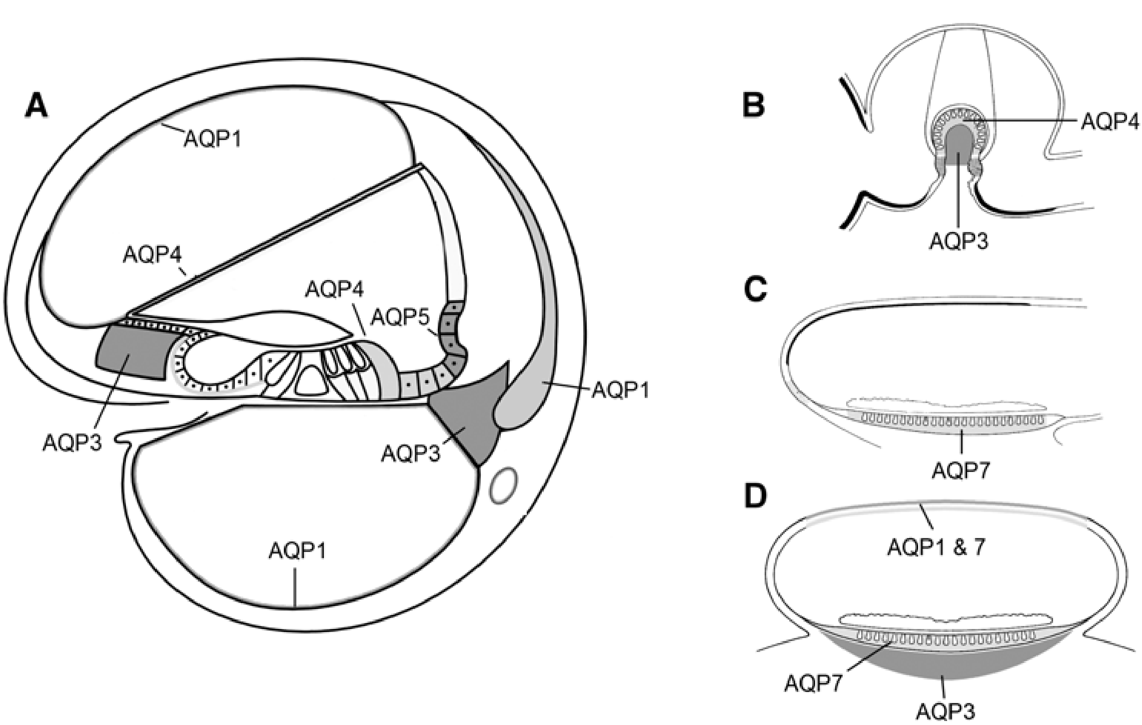


Figure 24–14 *A schematic diagram showing the localization of the expression of aquaporins (AQP1, AQP3, AQP4, AQP5, AQP7) in the different parts of the inner ear sensory receptors:* ***(A)*** *cochlear duct,* ***(B)*** *crista of a semicircular duct, and* ***(C,D)*** *maculae of the utricle and saccule. (Adapted from Huang D, Chen P, Chen S, Nagura M, Lim DJ, Lin X. Expression patterns of aquaporins in the inner ear: evidence for concerted actions of multiple types of aquaporins to facilitate water transport in the cochlea. Hear Res 2002;165:85–95. Reprinted with permission.)*

External Sulcus Cells (Root Cells)

External sulcus cells are found between Claudius' cells and the cells of the spiral prominence (**Fig. 24–2**). Their cell surfaces are slender and hexagonal shaped, but their cell bodies resemble the roots of a tree. Although the cell surfaces near the spiral prominence are facing the endolymph, Claudius' cells cover the rest of the surface of the external sulcus cells. These cells establish gap junctions with neighboring cells, including Claudius' cells and the type II fibrocytes of the spiral ligament. A unique feature of this cell type is the expression of solute carrier family 26, member 4 (SLC26A4) or pendrin (membrane chloride/solutes transporter) (**Fig. 24–15**). Pendrin gene mutations cause Pendred's syndrome and nonsyndromic hereditary hearing loss (DFNB4), which is characterized by high frequency, fluctuating, and sometimes progressive sensorineural hearing loss, and an enlarged vestibular aqueduct or Mondini's dysplasia (see Chapter 19). Frequencies of SLC26A4 mutations in East and South Asians are relatively high and account for ~5% of recessive deafness. The expression of AQP5 in the external sulcus cells and spiral prominence cells (**Fig. 24–14**) suggests that these cells are also involved in fluid regulation of the cochlea.

REISSNER'S MEMBRANE

Reissner's membrane divides the perilymph of scala vestibuli from the endolymph of the scala media and consists of two cell layers: epithelial cells that face the endolymph and mesothelial cells that face the perilymph (**Fig. 24–2**). The epithelial cells form tight junctions, providing an effective barrier between perilymph and endolymph, whereas the mesothelial cells are arranged loosely touching each other, allowing ions and macromolecules to freely pass through. Reissner's membrane can stretch extensively in Meniere's disease patients who have severe endolymphatic hydrops. This function is attributed to the presence of ion channels present on the cell membrane. Among the ion channels described are stretch-activated nonselective cationic channel (SA channel), voltage-sensitive cholide (Cl^-) channel, a K^+ channel, and Na^+, K^+–adenosinetriphosphatase (ATPase), present on the apical surface of the epithelial cells. The Reissner's membrane epithelial cells also express AQP4 and AQP7

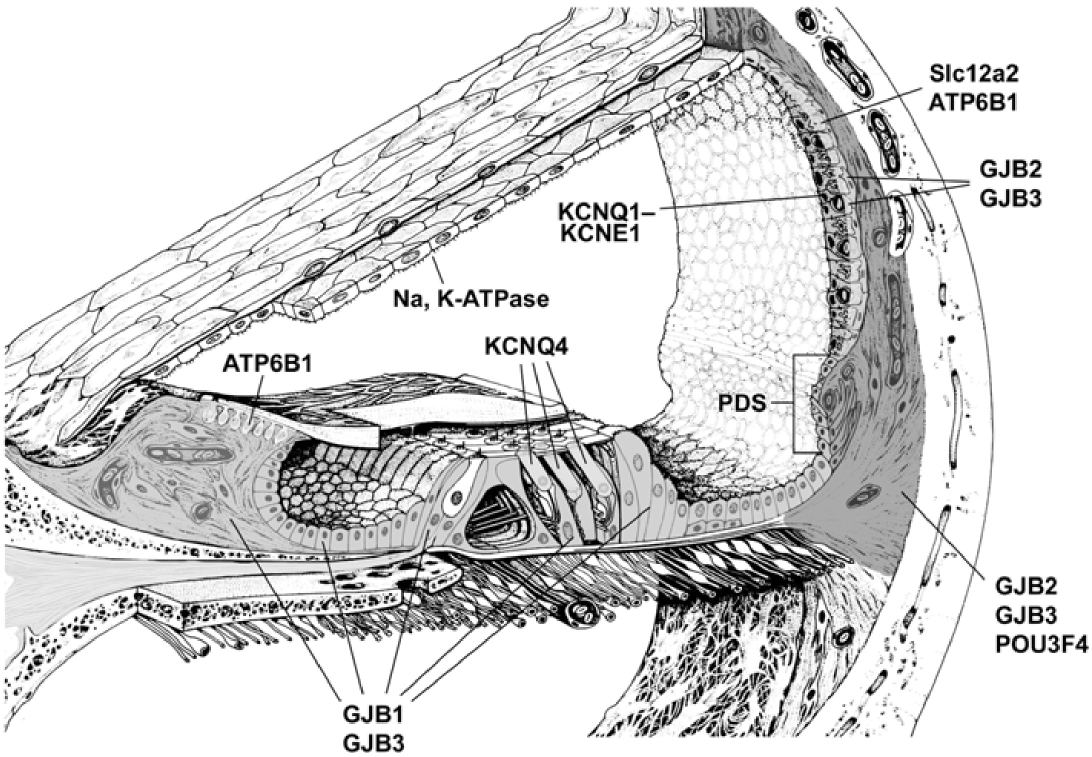


Figure 24–15 *The cochlea of the inner ear showing the location of key proteins involved in the recycling of K^+ ions. The transcription factor POU3F4 is expressed in the fibrocytes of the spiral ligament; KCNQ4 is expressed in the outer hair cells; connexins 26 (GJB2) and 31 (GJB3) are expressed in the supporting cells and fibrocytes of both the spiral ligament and the spiral limbus. (Adapted from Steel, K.P. Perspectives: biomedicine. The benefits of recycling. Science 1999; 285:1363–1364. Reprinted with permission.)*

(**Fig. 24–14**), suggesting that these cells play a role in fluid regulation of the cochlea.

CELLS AND STRUCTURES INVOLVED IN SOUND TRANSDUCTION

Tectorial Membrane

The tectorial membrane (TM) is an extracellular gelatinous membrane structure covering the reticular lamina; it makes direct contact with the tallest cilia of the OHC stereociliary bundles. The TM consists of distinct substructures: a fibrous main body with amorphous ground substance and several amorphous membrane structures: Hensen's stripe (which is located right above the inner hair cells), Kimura's membrane (formerly known as Hardesty's membrane), a cover net, and a marginal band or net (**Figs. 24–3, 24–4**).

Kimura's membrane forms the undersurface of the TM and is restricted to the area covering the OHCs. The TM is made up of a bundle of two morphologically distinct types of fibrils: Type A fibrils are long and straight, with a diameter of ~10 nm, and are mainly present in the fibrous layer, whereas type B fibrils are short and coiled filaments, ~15 to 20 nm in diameter, and are largely found in the amorphous portion of the TM. The fibrillar bundles are arranged radially and slightly slanted along the direction of the OHC stereociliary bundle angle of their W formation, suggesting the needed mechanical force it must exert for efficient ciliary deflection. Considerable evidence exists to demonstrate that OHC ciliary bundles are strongly coupled to the TM undersurface (Kimura's membrane), as evidenced by consistent imprinting of OHC stereociliary bundles on this membrane, whereas the stereociliary bundles of the IHC are not. Although there is evidence showing that the latter may have made contact with Hensen's stripe, few imprints of IHC stereociliary bundles on the edge of the Hensen's stripe have been observed. It is speculated that the stereociliary bundles of the IHCs are freestanding and that they are deflected by the endolymph drag produced by the shearing forces produced between the TM and the surface of the organ of Corti.

Biochemical studies have shown type II collagen, which constitutes 50% of the total protein of the TM, to be a major component; type IX and type V collagens were shown to be minor components of the TM. In situ hybridization for *COL2A1* messenger ribonucleic acid (mRNA) in human fetal tissue suggests that type II collagen is largely secreted by the inner ridge cells (the greater epithelial ridge) of Kölliker's organ during development. Proteoglycans containing keratan sulfate and chondroitin sulfate are believed to be the major components of the noncollagenous structures of the TM. It is now known that these noncollagenous structures are tectorins unique to the inner ear. High molecular weight tectorin may actually be a form of "light" keratan sulfate proteoglycan that is antigenically related to mucin. In mucociliary transport systems, the motile cilia of the ciliated cells must couple to the overlying mucous blanket (see Chapter 39). Sequence analysis of mouse tectorin has shown similarity to the components of the sperm–egg adhesion system. It appears that tectorin is needed to securely adhere the tips of stereocilia (particularly those of the OHCs) to the tectorial membrane. Targeted deletion of the extracellular matrix molecule α-tectorin in mouse resulted in a detached tectorial membrane phenotype; these animals show 35 dB reduction in hearing sensitivity and changes in gain and timing information. Mutations in *TECTA*, which encodes α-tectorin, result in nonsyndromic hereditary deafness (DFNA8 and DFNA12) (see Chapter 19). Another glycosylated protein, "otogelin," encoded by the gene *Otog*, is found under the surface of all gelatinous membrane structures of the inner ear, such as TM, otoconial membranes, and cupulae of vestibular sensory organs. In *Otog* knockout mice, both auditory and vestibular functions are impaired, and there is detachment of membranous structures in the vestibular sensory organs and disruption of the fibrillar networks in the TM of the organ of Corti. Another protein, "otoancorin," is localized to the limbal attachment zone of the TM, and mutations of the corresponding gene in humans, *OTOA*, results in nonsyndromic deafness (DFNB22). It appears that this protein is needed to anchor the TM to the spiral limbus. It is suggested that the tectorial membrane extracellular matrix molecules ensure that OHCs can respond effectively to basilar membrane motion and that feedback is delivered with the appropriate gain and timing required for amplification.

Basilar Membrane

The basilar membrane (BM) is composed mainly of an extracellular homogeneous material with numerous bundles of fibers (which consist of fine filaments) running radially along its width. The upper part of the BM is covered by basement membrane of the overlying epithelial cells. The spindle-shaped mesothelial cells (tympanic border cells) loosely cover the perilymphatic side of the BM, with their long cell bodies arranged along its length. Many bundles of the fibrils are often exposed directly to the perilymph, suggesting free communication of the perilymph with the internal structure of the BM. In adult animals, only a few connective tissue cells are found. Fibrillar bundles form either a single stratum in the arcuate zone (pars tecta) or a double stratum in the pectinate zone (pars pectinata) (**Figs. 24–3, 24–4**). The pars tecta is partly covered by the osseous spiral lamina; thus it is immobile. The pars pectinata is the mobile part of the BM. These filaments of BM fibrils merge with the nondirectionally arranged filaments of the spiral limbus in the medial part of the BM and with directionally arranged filaments of the spiral ligament in the lateral part of the BM. The filaments of the BM are made of types II, IV, and XI collagens. The major components of the noncollagenous ground substance are believed to be proteoglycan and fibronectin. Tenascin is also found in the ground substance of the BM.

LATERAL WALL TISSUES

Stria Vascularis

The lateral wall of the cochlea consists of the stria vascularis, the spiral prominence, and the spiral ligament (**Fig. 24–2**). The stria vascularis consists of three cell layers: marginal cells, intermediate cells, and basal cells. The marginal cell is an epithelial cell with an extensive interdigitating basolateral surface, and it makes direct contact with blood vessels outside the basement membrane. Stria vascularis epithelial cells are the only epithelial cells that are in direct contact with blood vessels in the body. The intermediate cells are derived from migrating melanocytes of the neural crest during the development of the stria vascularis. The stria vascularis plays a crucial role in generating the positive endocochlear potential (EP) of the scala media, as well as in maintaining a high concentration of K^+ of the endolymph. A mutant mouse with a neural crest melanocyte migration deficiency lacks intermediate strial cells and has no EP. The characterization of K^+ ion channels and ion transporters, as well as the connecting cellular gap-junction apparatus needed for K^+ ion recycling, has provided new insights into the mechanisms of maintaining cochlear homeostasis. Mutations of ion channels and gap-junction proteins have been shown to result in deafness and fluid imbalance in both humans and in animals in which these proteins have been either mutated or knocked out. Moreover, these observations have brought a heightened awareness of the complexity and specificity of the various inner ear cell types and their roles in the maintenance of the intricate and exquisite regulation of inner ear homeostasis.

Spiral Ligament

The spiral ligament provides support for the attachment of the basilar membrane, stria vascularis, and Reissner's membrane (**Fig. 24–2**). The spongy-appearing spiral ligament is made up of a meshwork largely of type I collagen secreted by spiral ligament fibrocytes, and its structure is bathed with perilymph. The extracellular matrix protein cochlin, encoded by *coagulation factor C homolog (COCH)* gene, is expressed in the spiral ligament and spiral limbus and is believed to be secreted by the fibrocytes. *COCH* mutation causes nonsyndromic dominant DFNA9 deafness with vestibular malfunctions similar to Meniere's disease (because of fluctuating hearing loss and tinnitus) (see Chapter 19).

There are five distinct types of spiral ligament cells (types I–V) that have been described based on carbonic anhydrase, Na^+K^+-ATPase and creatine phosphokinase-BB isozyme (CPK-BB) patterns of expression (**Table 24–1**). These cells are also distributed in clusters at discrete locations within the spiral ligament (**Fig. 24–16**). Type V cells express AQP1, and type IV cells express AQP3. The exact role(s) that these aquaporins play in inner ear homeostasis is/are not yet known. A more detailed discussion on inner ear aquaporins is presented later in this chapter.

Spiral Prominence Cells

Spiral prominence cells have flat to cuboidal cell bodies with a hexagonal luminar cell surface. Spiral prominence cells are segregated from the stria vascularis marginal cells by spindle-shaped border cells. The exact physiological role of this cell in the cochlea is not yet fully elucidated, but because they express SLC26A4 (pendrin) (**Fig. 24–15**), AQP5 (**Fig. 24–14**), and ENaC, their functions are suggested to be associated with the ion and fluid regulation of endolymph.

Spiral Limbus and Interdental Cells (Huschke's Teeth Cells)

Spiral limbus is made up of spiral ligament-like connective tissue and epithelial cells covering the surface of the spiral limbus known as interdental cells (IDCs). IDCs form well-arranged rows, making furrows of connective tissues known as Huschke's teeth. Spiral ligament fibrocytes belong to type II cells of the spiral ligament (see **Table 24–1**, **Fig. 24–16**) and express AQP3 (**Fig. 24–14**). As in the spiral ligament, the extracellular matrix protein cochlin is found in the spiral prominence. IDCs have a goblet-shaped cell body with a hexagonal

Table 24–1 Classification of connective tissue fibrocyte subtypes by inner ear regions

Region	Cell Types	Kv3	CA II	CA III	Na^+, K^+-ATPase	CPKBB	NKCC
Spiral ligament	Type I	+	+ to + + +	+ to + + +	−	− to + +	−
	Type II	−	−	−	+ + +	−	+
	Type III	+	+ to + + +	+ to + + +	−	+ to + + +	−
	Type IV	+	− to + + +	− to + + +	+ + +	− to + +[a]	−
	Type V	−	−	−	+	−	+
	Suprastrial area (superficial)		−	−	+ + +	−	
	Suprastrial area (deep)		− to + + +	− to + + +	−	− to + + +	
Spiral limbus	Spiral limbus proper	+	−	−	−	−	−
	Supralimbal area		−	−	+ + +	−	
Vestibular system	Under sensory epithelium		+ to + + +	−[b]	− to + +	− to + +[a]	

[a]A minority of the cells appeared stained.
[b]Except for clusters of stained fibrocytes under transitional (marginal) cells in the sacculus.
ATPase, adenosinetriphosphatase; CPKBB, creatine phosphokinase-BB; voltage-gated K^+ channel, K3; carbonic anhydrase, CA II, III; sodium–potassium–chloride co-transporter, NKCC
Adapted from Spicer SS, Schulte BA. Differentiation of inner ear fibrocytes according to their ion transport related activity. Hear Res 1991;56:53–64.

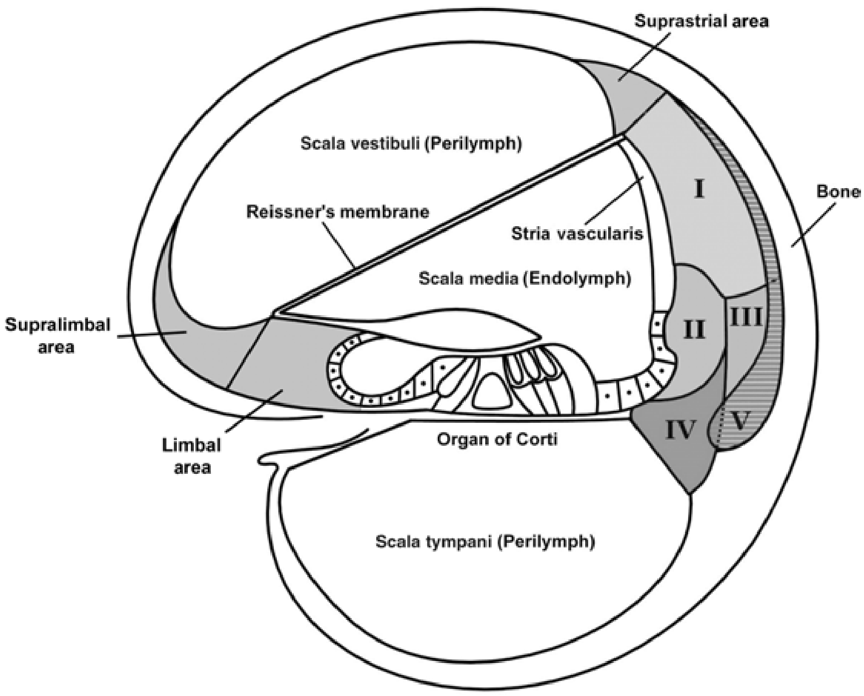


FIGURE 24–16 A diagrammatic presentation of the distribution of subtypes of spiral ligament cells (types I, II, III, IV, and V). Fibrocytes in the spiral limbus are tentatively labeled as type II cells. (Courtesy of D.J. Lim.)

umbrella-like phalanx to which the tectorial membrane attaches. The phalanx of the IDC sometimes invaginates into the cell body to form a cavity lined by microvilli, suggesting active fluid transport by this cell (**Fig. 24–17**). Because of the close attachment of the TM to these cells during development as well as in the adult cochlea, it is generally believed that this cell is also involved in both the production and the maintenance of the TM. It is also suggested that IDCs are responsible in the production of the glycosaminoglycans and/or tectorins of the TM. Experimental evidence suggests that pilocarpin can elicit secretory activity by the IDCs. Several ion channels are known to be expressed in the IDC, such as voltage-gated K^+ channel (Kv31b), NKCC1 K^+ ion co-transporter, AE2 anion exchanger, carbonic anhydrase, and the ATPase-gated ion channel P2X2, suggesting that the IDC plays an active role in the ion and fluid regulation of the endolymph. Recent experimental evidence suggests that IDC cells may be involved in K^+ recycling from the IHCs, as evidenced by the degeneration of the cells involved in the proposed medial K^+ recycling pathways.

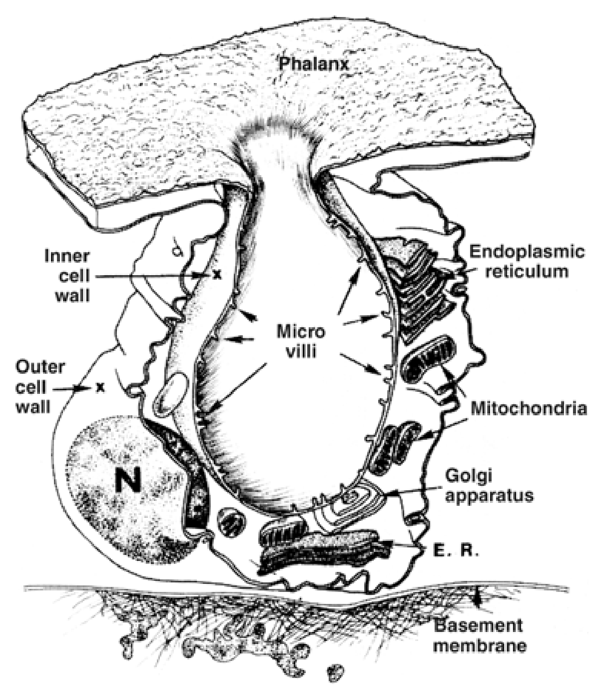


Figure 24–17 An artistic rendering of an interdental cell with a cytoplasmic duct. The tectorial membrane is not shown in this drawing. Also, neighboring interdental cells are not drawn; therefore, the basement membrane covers only part of this cell. ER, endoplasmic reticulum; N, nucleus. (From Lim, D.J. Morphology and function of the interdental cell—an ultrastructural observation. J Laryngol Otol 1970;84:1241–1256. Reprinted with permission. Drawing by Nancy Sally.)

INNER EAR HOMEOSTASIS AND PROPOSED POTASSIUM (K^+) RECYCLING PATHWAYS

The inner ear is a fluid-filled sensory organ that requires a precisely controlled ionic environment, oxygen level, water balance, and pH for its functions and efficient machinery to remove metabolites. The cochlea has three compartments: scala media, scala vestibuli, and scala tympani (**Fig. 24–1**). The self-contained scala media compartment is filled with K^+-rich endolymph, a unique extracellular fluid, with an ion composition unlike that found anywhere else in the body. The scala tympani and the scala vestibuli are filled with Na^+-rich perilymph that is a typical extracellular fluid with an ionic composition comparable to plasma or cerebrospinal fluid (CSF). These latter two scalae are connected together through the helicotrema and to the CSF through the cochlear aqueduct of the scala tympani.

The low concentration of Na^+ in the endolymph is thought to be maintained by the epithelial sodium (Na) channel (ENaC). Cells of the stria vascularis, spiral prominence, spiral limbus, spiral ligament, and spiral ganglion all express the subunits of ENaC.

SLC26A4 (pendrin) encoded by the *PDS* gene is a transmembrane anion/base exchanger expressed in specific cell types believed to be involved in ion and fluid regulation of the endolymph, such as apical membrane of cells of the spiral prominence and outer sulcus cells with their root processes, transitional cells of the vestibular sensory organs, and the apical membranes of cells that compose the endolymphatic sac. SLC26A4 (*PDS*) mutant (pendrin-deficient) mice are deaf and showed pathology of the stria vascularis and near zero endocochlear potential, but they have a near normal potassium concentration in their endolymph. Thus it is suggested by this observation that pendrin serves a key role in the functioning of the basal and/or intermediate cells of the stria vascularis to maintain the endocochlear potential, but not potassium secretion by the marginal cells. The positive endocochlear potential seen in the scala media of the cochlea is believed to be generated by the intermediate cells and basal cells of the stria vascularis. *PDS* gene mutation in humans causes Pendred's syndrome or nonsyndromic sensorineural hearing loss DFNB4 (see Chapter 19). Both Pendred's syndrome and DFNB4 phenotypes are associated with enlargement of the vestibular aqueduct. Lack of pendrin expression leads to deafness and expansion of the endolymphatic compartment in the inner ears of *Foxi1* (winged helix/forkhead gene) null mutant mice, thus suggesting that *Foxi1* is an upstream regulator of pendrin and that the phenotype seen in *Foxi1* null mice is, at least in part, due to defective pendrin-mediated chloride ion resorption by endolymphatic duct/sac epithelium.

Ion Channels Involved in Proposed K^+ Recycling Pathways

The maintenance of ion homeostasis of the inner ear is essential to this organ's normal function. As mentioned earlier, activation of hair cells by acoustic stimuli causes an influx of K^+ ions from the endolymph into the sensory hair cells via apical transduction channels. High levels of potassium, if allowed to accumulate in the sensory cells, are toxic to these cells and must be removed from them. The K^+ ion in the sensory cell is rapidly removed from the hair cells via the K^+ ion channel KCNQ4, expressed on the basolateral membrane of these sensory cells (**Figs. 24–2, 24–15**).

Proposed Lateral K^+ Recycling Pathways from OHCs

The potassium released into the extracellular space surrounding the OHCs then enters the adjacent supporting cells, such as Deiters' cells and Hensen's cells (**Figs. 24–15, 24–18**). Once inside these cells, the K^+ ions begin to diffuse via the epithelial cell gap-junction system of Deiters' cells, Hensen's cell, Claudius' cell, and external sulcus cell (root cell), toward the lower part of the spiral ligament. The K^+ ions that are released into the extracellular space of the spiral ligament by root cells are then taken up by type II fibrocytes that are present adjacent to these root cells (**Fig. 24–18**). Both type I and type II fibrocytes express the $Na^+/2Cl^-/K^+$ co-transporter SLC12A2 and the Na^+/K^+ ATPases (ATP1A1/ATP1B1 and ATP1A2/ATP1B1 isoenzymes) that are responsible for K^+ absorption. Type II fibrocytes make contact via gap junctions with type I fibrocytes, which are found in the spiral ligament located behind the stria vascularis. Type I fibrocytes establish gap junctions with the basal cells of the stria vascularis, which in turn also establish gap junctions with the intermediate cells of the stria vascularis (**Fig. 24–18**). The K^+ ions move through the gap-junction system and are released into the intrastrial space via the KCNJ10 (Kir4.1) channels of the intermediate cells (**Fig. 24–19**). It is currently believed that the K^+ channel KCNJ10 is responsible for the generation of the endocochlear potential. Mice lacking the K^+ channel KCNJ10 in their intermediate cells do not generate an EP.

The K^+ released into the intrastrial space is then taken up by the strial marginal cells via the $Na^+/2Cl^-/K^+$ co-transporter (SLC12A2) and the Na^+/K^+ ATPases [ATP1A1/ATP1B1 ($\alpha 1$, $\beta 1$) and ATP1A2/ATP1B1 ($\alpha 2$, $\beta 1$)]. Strial marginal cells are not connected to intermediate cells via gap junctions. The strial marginal cells release K^+ into the endolymph via the K^+ channels (KCNQ1/KCNE1) that are expressed on their apical membrane (**Fig. 24–19**). This secretion of K^+ is required not only for maintaining the EP and providing the charge carrier for the transduction mechanism, but for maintaining a constant volume of

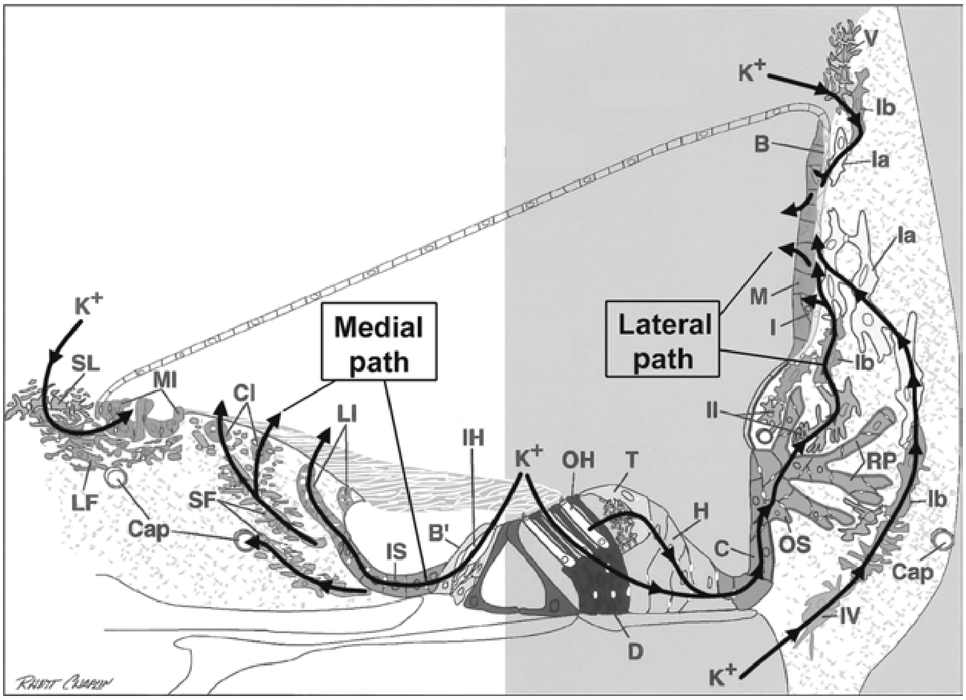


__Figure 24–18__ A schematic representation of the proposed medial and lateral transcellular routes for dispersal and conservation of K^+ effluxes from inner and outer hair cells during auditory transduction. Ions released from inner hair cells (IH) diffuse medially through border cells (B′), inner sulcus cells (IS), and lateral interdental cell (LI) columns to the undersurface of the tectorial membrane and from IS through stellate fibrocytes (SF) to capillaries (Cap) or to central interdental cells (CI) and the scala media. K^+ pumped from scala vestibuli into supralimbal cells (SL) flows down gradient to light fibrocytes (LF) and medial interdental cells (MI) for return to scala media. In the lateral route, K^+ effluxing from outer hair cells (OH) is resorbed by Deiters' cells (D) and tectal cells (T) and flows via gap junctions through Hensen's (H), Claudius' (C), and outer sulcus cells (OS) and their root processes (RP) to efflux into stroma maintained at a low K^+ level by the Na, K-ATPase activity of type II fibrocytes. K^+ subsequently diffuses via gap junctions through type I fibrocytes (Ia, Ib) and strial basal (B) and intermediate cells (I) into the intrastrial compartment kept low in K^+ by the pumping activity of strial marginal cells (M). K^+ resorbed by type V fibrocytes from scala vestibuli diffuses downhill through Ib, then Ia fibrocytes, to the stria. B, basal cell; Ia, Ib, II, IV and V, types of lateral wall fibrocytes. (Adapted from Spicer, S.S, Schulte, B.A. Evidence for a medial K^+ recycling pathway from inner hair cells. Hear Res 1998;118:1–12. Reprinted with permission.)

endolymph in the scala media. Mice lacking the $Na^+/2Cl^-/K^+$ co-transporter SLC12A2, as well as those lacking the luminal surface K^+ channel KCNQ1 or the K^+ channel KCNE1 subunits of the K^+ channel KCNQ1/KCNE1 in strial marginal cells, are deaf and also have vestibular abnormalities.

Proposed Medial K+ Recycling Pathways from IHCs

Potassium also needs to be recycled from inner hair cells. Recent experimental results suggest that the potassium released by the IHCs travels medially through inner phalangeal cells, border cells, inner sulcus cells (ISCs) directly or by way of limbal fibrocytes via gap junctions, similar to the spiral ligament, and is finally released back into the endolymph via the interdental cells (**Fig. 24–18**). Moreover, other studies have shown the presence of connexin-26 in both the connective tissue fibrocytes and the IDCs.

CONNEXINS AND THE ROLE OF GAP JUNCTIONS IN INNER EAR HOMEOSTASIS

Gap junctions are transmembrane channels that connect the cytoplasm of two adjacent cells. Two hemichannels (connexons) in the cell membranes of two connecting cells form a complete gap junction. Each connexon consists of six connexins. Thirteen different connexin genes have been identified thus far. They are classified according to their molecular weight, leading to different types of channels. Because different gap junction channels have different molecular permeabilities, different connexins may permit different types of signaling between cells. Single gap junction

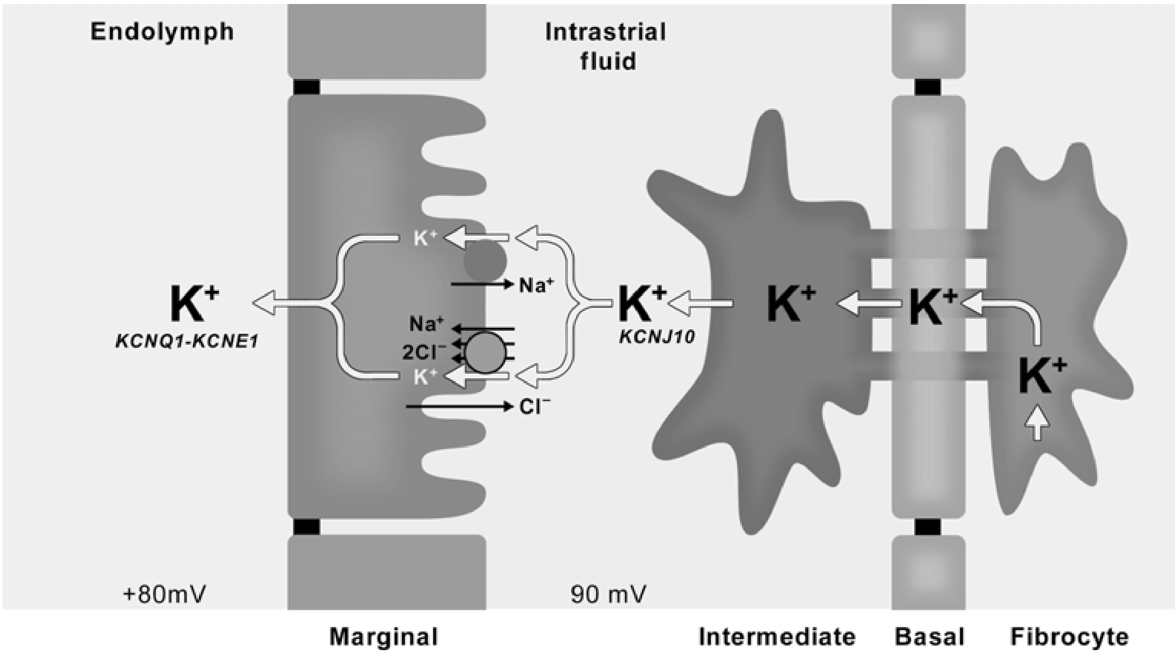


Figure 24–19 *An artist's conception of ion channels and gap junctions involved in potassium ion recycling via spiral ligament type I fibrocytes, and the basal, intermediate, and marginal cells of the stria vascularis. It is believed that cytoplasmic K^+ in the spiral ligament fibrocytes is transferred to basal cells and intermediate cells of the stria vascularis via gap junctions connecting these cells. Intermediate cells do not form gap junctions with marginal cells; thus the K^+ is released into the strial intercellular space via the KCNJ10 K^+ channels of the intermediate cells. Intercellular K^+ is then taken up by marginal cell Na^+, K^+ ATPase, and Na^+, $2Cl^-$, K^+ co-transporter expressed in the basolateral membrane and secreted to the endolymph via KCNQ1-KCNE1 K^+ channels expressed in the apical surfaces of the marginal cells of the stria vascularis. (Adapted from Wangemann, P. K^+ cycling and the endocochlear potential. Hear Res 2002;165:1–9. Reprinted with permission.)*

channels may be composed of the same connexon (homomeric) or two different connexons (heteromeric) channels. The main function of the gap junction is to provide direct intercellular communication pathways allowing rapid exchange of ions, second messengers, and metabolites up to ~1 kD molecule in size. Approximately 20 connexins have been identified within the human and mouse genomes.

To date, four connexins, including Cx26 (GBJ2), Cx30 (GBJ6), Cx31 (GBJ3), and Cx43 (GJA1), have been identified in the cochlea, and three connexins, including connexin 26, 30, and 43, have been identified in the vestibular organs. Recent data have shown that connexins 26 and 30 can form heteromeric connexons. Supporting cells also express connexin 43, although at much lower levels. Connexons are thought to be grouped into two systems—the epithelial cell gap junctions and the connective tissue gap junctions—with both systems believed to be contributing to the recycling of high concentrations of K^+ ions within the inner ear. The epithelial gap junctions are found between the supporting cells of the organ of Corti, cells lining the endolymph side of the basilar membrane, and also the basal cell and intermediate cells of the stria vascularis. All of these supporting cells are connected by an epithelial cell gap-junction system expressing GJB2 (Cx26), GJB3 (Cx30), and GJB6 (Cx31) (**Fig. 24–15**).

Mutations in four members of the connexin gene family, connexins 26, 30, 31, and 32, have been shown to underlie distinct genetic forms of deafness (see Chapter 19). Recessive mutations in the gene encoding connexin 26 are the most common cause of childhood-onset deafness (DFNB1). Indeed, connexin 26 mutations account for up to 50% of cases of nonsyndromic sensorineural hearing loss in some populations. More than 20 mutations in the *GJB2* (connexin 26) gene are associated with DFNB1, a prevalent type of autosomal recessive nonsyndromic sensorineural deafness. A frame shift mutation of a single guanine in position 35 (35delG) accounts for nearly 70% of connexin 26 mutations. Mutations in connexin 30 and 31 result in DFNA3 and DFNA2, respectively, and mutations in connexin 32 in patients with X-linked Charcot-Marie-Tooth disease that also causes sensorineural deafness. In addition, mutations in connexin 43 are associated with nonsyndromic autosomal recessive deafness. The role of connexin mutations in the pathophysiology of deafness opens new opportunities in the molecular diagnosis of hearing loss and eventually may lead to the development of novel therapeutic strategies for the treatment of some forms of deafness.

There is no question as to the importance of gap junctions in inner ear homeostasis, which is critical for hearing. The proposed role of gap junctions in K^+ recycling is an attractive hypothesis. However, many unresolved questions remain, such as to how the unidirectional flow of K^+ is maintained. Another possible function of the gap junction system in the cochlea could be the removal of metabolites and/or delivery of nutrients from the capillary-rich area to the sensory cells, where there are no blood capillaries nearby except within the developing ear. They may also play a role in the differentiation of the cells during development.

REGULATION OF pH IN THE INNER EAR

The mechanisms that regulate endolymphatic pH are not well understood. Because the pH of endolymph is close to that of blood, it is postulated that H^+ ions are being secreted into the endolymph. The intracellular pH of nearly all mammalian cells is controlled by the Na^+/H^+ exchanger (NHE). It has been demonstrated that cochlear cells, including those of the stria vascularis and the OHCs, also regulate pH via NHE. The pH-regulating proteins vH^+ ATPase (in the apical membrane) and Cl^-/HCO_3 vascular H^+ (in the basolateral membrane) were localized in IDCs and the epithelial cells of the endolymphatic sac. Cochlear cell types with diffuse cytoplasmic staining of vH^+ ATPase and a basolateral localized Cl^-/HCO_3 exchanger were observed in the inner hair cells, root cells, and a subset of supporting cells in the organ of Corti. Hair cells of the vestibular sensory organs express these proteins. A vH^+-ATPase (ATP6B1) mutation is known to cause recessive nonsyndromic hearing loss. This protein is expressed in IDCs and within endolymphatic duct epithelial cells. Mice lacking Atp6b1 did not show any evidence of hearing loss, suggesting that other proton-transporting mechanisms or pH-buffering systems must allow the mouse inner ear to compensate for lack of normal ATP6B1 activity.

There is good evidence that several inner ear cells, such as spiral ligament fibrocytes, external sulcus cells (root cells), strial cells, and endolymphatic sac epithelial cells, express carbonic anhydrase (CA) subtypes. CA catalyzes the rapid interconversion of carbon dioxide and water into carbonic acid, protons, and bicarbonate ions, and it may be involved in the maintenance of the pH of the endolymph.

There was a recent demonstration of acid-sensing ion channel 2a (ASIC2a) expression in the cochlear supporting cells, inner and external sulcus cells, and spiral ganglion cells. While the role of ASIC2a in the inner ear is not yet clear, it is most likely involved in the detection of pH changes that could occur upon cell injury. In certain cases, the ASIC can also function as a mechanosensor. Whether or not ASIC is the postulated pressure sensor in the inner ear that may be involved in the feedback mechanism for fluid regulation remains to be determined.

INNER EAR AQUAPORINS (WATER TRANSPORTERS)

The aquaporins (AQPs) are a family of small transmembrane water transporters that play a role in regulating homeostasis of inner ear fluids. Multiple AQPs are expressed in the inner ear: AQP1, AQP2, AQP3, AQP4, AQP5, AQP6, AQP7, and AQP9 (**Fig. 24–14**). Almost all of these aquaporins are expressed in both the cochlea and the endolymphatic sac (ES), with the exceptions being that AQP5 is expressed in the cochlea but not in the ES, and AQP2 and AQP6 is expressed in the ES but not in the cochlea. Neither AQP5 nor AQP6 appears to be expressed in the vestibule. Because the maintenance of endolymph homeostasis is critical for the inner ear to perform its functions of hearing and maintaining balance, the dysregulation of AQPs may contribute to inner ear homeostasis disorders that include Meniere's disease and endolymphatic hydrops. Subtypes of aquaporins are expressed in specific cell types and in the subtypes of spiral ligament cells and fibrocytes of inner ear connective tissue (**Fig. 24–14**). Such a distinct distribution of the aquaporins suggests regionalized control of fluid homeostasis within the inner ear.

In transgenic null mice lacking AQP4, the auditory brainstem response (ABR) thresholds were increased, indicating a hearing loss. Thus it is likely that AQP4 plays a role in hearing by facilitating rapid osmotic equilibration in epithelial cells of the organ of Corti, particularly Hensen's cells and Claudius' cells, Reissner's membrane, and the connective tissue cells of the crista ampullaris (**Fig. 24–14**), which can all be subjected to large K^+ fluxes during mechanoelectric signal transduction. In contrast, the ABR thresholds were not affected by AQP1, AQP3, or AQP5 deletion mutations. The effect of a null mutation in AQP2 has not been studied because the mice die of diabetes insipidus in the first week after birth. Intratympanic injection of steroids upregulates AQP1 expression in a dose-dependent manner, suggesting a mechanism whereby steroids may affect water homeostasis in the inner ear. AQP1 is largely expressed in the type IV fibrocytes of the spiral ligament (**Fig. 24–16**).

In the principal cells of the kidney, the levels of AQP2 and its localization in the apical membrane of these cells are regulated by the antidiuretic hormone vasopressin, and over- or underexpression of AQP2 leads to several different disease states. In addition, the long-term regulation of AQP3 appears to be influenced by vasopressin. In the inner ear, AQP2 is exclusively expressed in the endolymphatic sac. Evidence to support an important role for the regulation of fluid homeostasis by the vasopressin–AQP2 system in the inner ear comes from both clinical and experimental studies. Patients with Meniere's disease have elevated levels of plasma vasopressin. Acute and chronic application of vasopressin to rats and guinea pigs produces endolymphatic hydrops within the cochlea. In addition, acute vasopressin treatment causes an upregulation of AQP2 mRNA expression in both the ES and the cochlea. The most compelling evidence that endolymphatic fluid homeostasis in the inner ear may be regulated by the vasopressin–AQP2 system comes from a very recent study in which treatment of rats with a vasopressin-2 receptor antagonist resulted in a reduction in experimentally induced endolymphatic hydrops and in AQP2 mRNA levels. Although the exact target of the vasopressin-2 receptor antagonist in the inner ear is not known, nor is it clear if the cellular regulation of AQP2 by vasopressin in the inner ear differs from that in the kidney, these results suggest that AQP2 may play an important role in the regulation of endolymph volume and that dysregulation of AQP2 may contribute to the development of Meniere's disease and endolymphatic hydrops.

SUMMARY

There is a highly elaborate and complex network of inner ear homeostasis systems operating within the cochlea. How they are interrelated with one another and elucidation of how their feedback loops and signaling pathways are regulated will ultimately help to resolve many puzzles of cochlear homeostasis and how its dysfunction results in sensorineural hearing loss and many other types of inner ear disorders. Many new drug targets for inner ear homeostasis disorders will open new chapters for the diagnosis and management of the inner ear disorders for which there are no known cures.

Suggested Readings

Ahituv N, Avraham KB. Mouse models for human deafness: current tools for new fashions. Trends Mol Med 2002;8:447–451

Beitz E, Zenner HP, Schultz JE. Aquaporin-mediated fluid regulation in the inner ear. Cell Mol Neurobiol 2003; 23:315–329

Huang D, Chen P, Chen S, Nagura M, Lim DJ, Lin X. Expression patterns of aquaporins in the inner ear: evidence for concerted actions of multiple types of aquaporins to facilitate water transport in the cochlea. Hear Res 2002;165:85–95

Ikeda K. Gene-based deafness research: ion transport and hearing. Tohoku J Exp Med 2004;202:1–11

Kachar B, Parakkal M, Kurc M, Zhao Y, Gillespie PG. High-resolution structure of hair-cell tip links. Proc Natl Acad Sci USA 2000;97:13336–13341

Kalinec F, Holley MC, Iwasa KH, Lim DJ, Kachar B. A membrane-based force generation mechanism in auditory sensory cells. Proc Natl Acad Sci USA 1992;89:8671–8675

Kalinec F, Kachar B. Structure of the electromechanical transduction mechanism in mammalian outer hair cells. In: Flock A, Ottoson D, Ulfendahl M, eds. Active Hearing. Amsterdam: Elsevier Science; 1995:179–191

Kikuchi T, Kimura RS, Paul DL, Takasaka T, Adams JC. Gap junction systems in the mammalian cochlea. Brain Res Brain Res Rev 2000;32:163–166

Lim DJ. Morphology and function of the interdental cell—an ultrastructural observation. J Laryngol Otol 1970;84:1241–1256

Lim DJ. Functional structure of the organ of Corti: a review. Hear Res 1986;22:117–146

Lim DJ. Scanning electron microscopic morphology of the ear. In: Paparella MM, Shumrick DA, Gluckman JL, Meyerhoff WL, eds. Otolaryngology, Basic Sciences and Related Principles. Philadelphia: WB Saunders; 1991:459–494

Lim DJ, Kalinec F. Cell and molecular basis of hearing. Kidney Int Suppl 1998;65:S104–S113

Morton CC. Genetics, genomics and gene discovery in the auditory system. Hum Mol Genet 2002;11:1229–1240

Rabionet R, Lopez-Bigas N, Arbones ML, Estivill X. Connexin mutations in hearing loss, dermatological and neurological disorders. Trends Mol Med 2002;8:205–212

Raphael Y, Altschuler RA. Structure and innervation of the cochlea. Brain Res Bull 2003;60:397–422

Santi PA. Cochlear microanatomy and ultrastructure. In: Jahn AF, Santos-Sacchi J, eds. Physiology of the Ear. New York: Raven Press; 1988:173–199

Schneider ME, Belyantseva IA, Azevedo RB, Kachar B. Rapid renewal of auditory hair bundles. Nature 2002;418:837–838

Slepecky NB. Structure of the mammalian cochlea. In: Dallos P, Popper AN, Fay RR, eds. The Cochlea. New York: Springer-Verlag; 1996:44–129

Spicer SS, Schulte BA. Differentiation of inner ear fibrocytes according to their ion transport related activity. Hear Res 1991;56:53–64

Spicer SS, Schulte BA. Evidence for a medial K+ recycling pathway from inner hair cells. Hear Res 1998;118:1–12

Steel KP. Perspectives: biomedicine. The benefits of recycling. Science 1999;285:1363–1364

Steel KP, Kros CJ. A genetic approach to understanding auditory function. Nat Genet 2001;27:143–149

Wangemann P. K^+ cycling and the endocochlear potential. Hear Res 2002;165:1–9

Wangemann P, Schacht J. Homeostatic mechanisms in the cochlea. In: Dallos P, Popper AN, Fay RR, eds. The Cochlea. New York: Springer-Verlag; 1996:130–185

Weber PC, Cunningham CD III, Schulte BA. Potassium recycling pathways in the human cochlea. Laryngoscope 2001; 111:1156–1165

Willems PJ. Genetic causes of hearing loss. N Engl J Med 2000;342:1101–1109

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. The basilar and tectorial membranes of the cochlea
   - A. Are of a uniform width but vary in thickness within the cochlea
   - B. Are of a constant width and thickness throughout the cochlea
   - C. Vary in both thickness and width in a base-to-apex pattern within the cochlea
   - D. Are middle ear structures and are not located within the cochlea
2. The stereocilia of the cochlea's auditory hair cells
   - A. Are true cilia and therefore contain a 9 + 2 arrangement of microtubules
   - B. Contain actin but without any high degree of molecular organization
   - C. Are composed predominantly of myosin molecules
   - D. Contain a highly ordered paracrystalline arrangement of actin filaments
3. The cell body of the outer hair cell
   - A. Does not have any distinguishing characteristics and appears much like that of the inner hair cell
   - B. Contains many mitochondria that are uniformly distributed throughout its cytoplasm
   - C. Contains a well-organized modification of its endoplasmic reticulum known as the subsurface cisternae that is located along the lateral wall of its lateral membrane
   - D. Is contacted predominantly by afferent nerve buttons from the type I auditory neurons
4. Reissner's membrane
   - A. Is composed only of type I fibrocytes and is located within the scala tympani
   - B. Divides the perilymph of the scala vestibuli from the endolymph of the scala media and consists of two cell layers (i.e., epithelial cells facing the endolymph in the scala media and a loosely organized layer of mesothelial cells facing the perilymph of the scala vestibuli)
   - C. Divides the perilymph of the scala vestibuli from the endolymph of the scala media but is an acellular structure composed of several extracellular matrix molecules
   - D. Divides the endolymph of the scala media from the perilymph of the scala tympani and consists of two cell layers (i.e., epithelial cells facing the endolymph in the scala media and a loosely organized layer of mesothelial cells facing the perilymph of the scala tympani)
5. The stria vascularis is responsible for maintaining the unique composition of the endolymph and the endolymphatic potential and is composed of
   - A. Three types of cells (i.e., marginal cells, intermediate cells, and basal cells)
   - B. A single cell type (i.e., marginal cells of epithelial origin)
   - C. Three types of cells (i.e., types I, II, and III fibrocytes)
   - D. Two different cell types but with none of the abovementioned cell types present

Chapter 25
HAIR CELL FUNCTION

PETER G. GILLESPIE

Hair cells, the sensory cells of the auditory and vestibular systems, convert sound and movements of the head into electrical signals. Hair cell integrity is crucial for hearing and balance; accordingly, many perturbations of hair cells lead to deafness or vestibular dysfunction.

Hair cells have two principal functions: they transduce auditory and vestibular stimuli, and they amplify low-intensity signals so that they can be more easily detected. This chapter will focus on the cellular features required for mechanoelectrical transduction and active amplification.

The structure of a hair cell gives important clues as to its function. Key features include the hair bundle, synaptic region, and, in the case of the outer hair cell, lateral wall. Once hair cell structure is understood, the mechanism of excitation of a hair cell can be revealed; movements of the auditory or vestibular partition lead to excitatory (or inhibitory) bundle movements. Excitation in turn initiates mechanoelectrical transduction, which is distinguished by its extremely fast time course. Low-intensity sounds are amplified at a specific frequency by the cochlear amplifier, probably residing in outer hair cells. Finally, excitation in a hair cell leads to synaptic release at specific sites; in turn efferent stimulation modulates the excitability of a hair cell.

HAIR CELL STRUCTURE

Hair cells are not true neurons but instead are considered neuroepithelial cells. Besides possessing features found in many epithelial cells, hair cells have specialized features required for transduction and synaptic transmission. Familiarity with epithelial and neuronal cells makes understanding hair cell structure quite simple (**Fig. 25–1**).

Perhaps the most important element of a hair cell is its hair bundle. Responsible for mechanical transduction, the hair bundle contains a cluster of 30 to 300 actin-based cellular processes called stereocilia. Containing several hundred actin filaments, each stereocilium is covered by the plasma membrane, so that the inside of a stereocilium is continuous with the inside of the cell. Stereocilia resemble the microvilli found on surfaces of many epithelial cells, although stereocilia are much wider and far longer.

Stereociliary actin filaments are heavily cross-linked together by the proteins fimbrin and espin. The cross-linked actin filaments endow the stereocilia with high rigidity. The number of actin filaments decreases, however, as the stereocilium enters the apical surface of the cell; only one or two dozen filaments project into the soma. At this taper region, the stereocilia are very compliant, so that

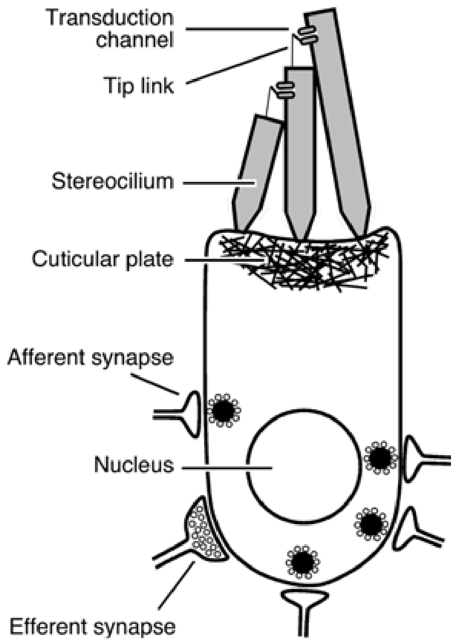


Figure 25–1 *Fundamental features of a generic hair cell. Transduction channels at stereocilia tips are gated by tip links; each stereocilium inserts into and is anchored by the cuticular plate. Hair cells have both afferent synapses (marked by their synaptic bodies) and efferent synapses.*

a stereocilium in isolation would bend at its base and remain straight along its length.

In all hair bundles, stereocilia are arranged in ascending ranks, so that the bundle resembles a pipe organ (cochlear hair cells) or a beveled needle (vestibular hair cells). The stereocilia are connected together by a variety of filamentous cross-links; some of these links may prevent bundle splaying, and one—the tip link—is intimately involved in hair cell transduction (below). Stereocilia are rigid, have compliant insertions, and are cross-linked by elastic linkages; accordingly, when the bundle is deflected, the entire bundle moves as a unit. Furthermore, a stereocilium slides along the adjacent stereocilium during a stimulus; this shearing action is thought to be critical for initiating mechanoelectrical transduction.

The hair bundle inserts into, and is anchored by, a meshwork of actin filaments, the cuticular plate. The cuticular plate extends across the apical surface of a hair cell, terminating just before another actin assembly, the circumferential actin band. These two actin-rich structures, in conjunction with intercellular junctions between hair cells and surrounding supporting cells, endow the apical surface of hair cell epithelia (called the reticular lamina in the cochlea) with considerable rigidity.

One other feature distinguishes hair cells, their synaptic region. Afferent and efferent synapses are found at the basal pole of cochlear hair cells and the basolateral surface of vestibular hair cells. Presynaptic structures of the afferent synapses are particularly evident; an electron-dense sphere, sometimes called the synaptic body, is surrounded by a halo of synaptic vesicles and marks the position of this synapse.

Outer hair cells have additional intracellular specialization, including their subsurface cisternae. Multiple layers of these membrane-bound intracellular organelles line the lateral wall of the hair cell. Their function is unknown. In addition, the cortical actin network is particularly well elaborated in outer hair cells, arranged so that changes in cell-surface area are converted into length changes of the cell.

Key Points

- Hair bundles consist of stereocilia.
- Stereocilia are made up of cross-linked actin filaments, covered by the plasma membrane.
- Hundreds of actin filaments arranged in parallel make the stereocilium stiff, except at the insertion into the apex of the cell.
- Elastic cross-links hold the bundle together, allowing the bundle to move as a unit.
- Stereocilia bend at their insertions and slide along each other.
- A hair bundle is anchored by the cuticular plate, an actin meshwork.
- Synapses are located at the basal pole or along the basolateral surface of a hair cell.
- Outer hair cells contain a well-organized cortical actin network and subsurface cisternae.

EXCITATION OF HAIR CELLS

Bending of the hair bundle, which allows sliding of adjacent stereocilia, is the key mechanical event during hair cell transduction. Force transmission of external signals to hair cells is thus a crucial element in auditory and vestibular function.

Sound energy is transmitted to the cochlea via the external auditory meatus, through the tympanic membrane and middle ear (allowing for impedance matching), and impinges on the oval window. Sound energy is transmitted through the cochlea to the round window, exciting the basilar membrane as it passes through.

Oscillating acoustic stimuli produce oscillation of the elastic basilar membrane. The elasticity of the basilar membrane varies along the cochlea; it is stiff and narrow at the base, where high-frequency sounds are encoded,

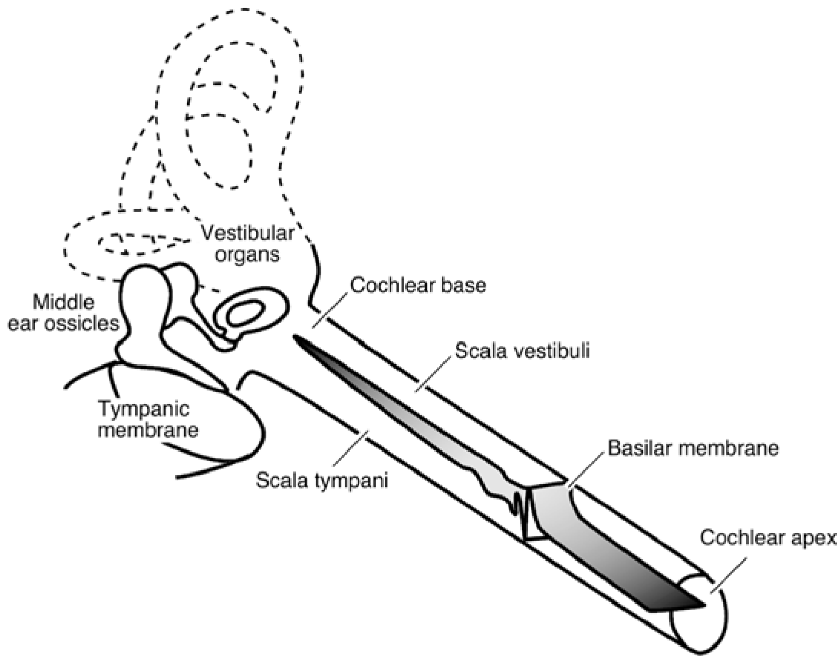


Figure 25–2 *Excitation of the basilar membrane. Sound–pressure oscillations are transmitted as vibrations of the middle ear ossicles. Vibration of the stapes footplate on the round window leads to vibrations of the basilar membrane, which peak at a characteristic point as the traveling wave of excitation moves from base (where high-frequency sounds are detected) to apex (where low-frequency sounds are detected).*

and compliant and wide at the apex, where low-frequency sounds are encoded. Acoustic stimuli move along the basilar membrane in a traveling wave, eventually encountering a region with sufficiently resonant properties that allows efficient displacement (**Fig. 25–2**).

The properties of the basilar membrane provide the first level of tuning within the cochlea. A high-frequency sound stimulus leads to a traveling wave that peaks at the base of the cochlea, near the oval window; no sound energy passes along the basilar membrane to regions that encode lower frequency sounds. A low-frequency sound stimulus, by contrast, travels through the region encoding high-frequency sounds but causes little or no displacement of the basilar membrane in this region. Only when the traveling wave reaches the region of the basilar membrane with the right mechanical properties does the sound stimulus lead to efficient basilar membrane displacement.

Acoustic stimulation thus leads to an up-and-down motion of the basilar membrane. How is this motion converted to bundle displacement? Hair cells are held rigidly laterally by the structural elements of the reticular lamina, the apical surfaces of the hair, and supporting cells. Tight and adhering junctions are particularly prominent in this region, endowing the reticular lamina with substantial lateral stiffness. Hair bundles are anchored (by hair cell somas) in the reticular lamina and, at least in the case of the outer hair cells, the tectorial membrane overlying them. The tectorial membrane, like the reticular lamina, is anchored in the spiral limbus. Up or down movement of the basilar membrane displaces the reticular lamina without moving the tectorial membrane; this stimulus leads to a shear stimulus of the hair bundle, and the bundle is moved in the positive or negative directions, along the axis of mechanical sensitivity (**Fig. 25–3**).

Inner hair cells are not thought to be coupled to the tectorial membrane and instead may be displaced by hydrodynamic forces arising during basilar membrane displacement. Unanchored hair bundles would seem to have a big disadvantage—Brownian motion due to water molecules banging on the bundle should be substantially greater than that motion elicited from an anchored hair bundle. Increased Brownian motion may, however, enhance hair cell sensitivity through stochastic resonance, a well-studied phenomenon whereby broad-band noise can enhance detection of small signals.

Vestibular hair bundles are excited by relative movement between the overlying mechanical structure, such as the otolithic membrane or cupula, and the apical surface of the vestibular epithelium. These mechanical structures are, in turn, coupled appropriately so that the hair cells detect head movements with the appropriate features.

In the utricle and saccule, the overlying otolithic membrane is coupled to a mass of otoconia, small calcium carbonate crystals of higher density than water. During a head acceleration along the horizontal (utricle) or vertical (saccule) planes, inertial forces ensure that the otoconia lag behind other components of the organ. Because the otoconia move less far than the organ itself, and because the otoconia are mechanically coupled to

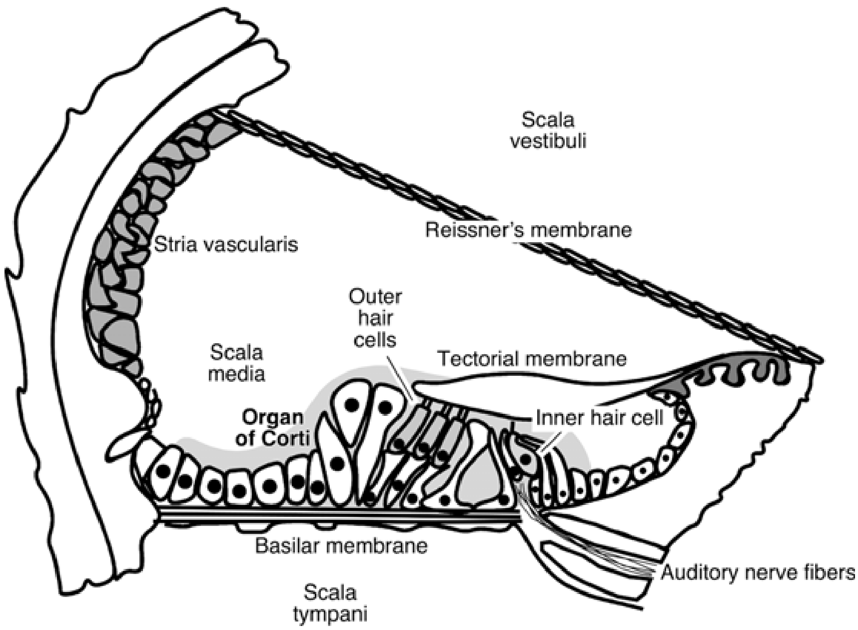


Figure 25–3 *Location of hair cells within the organ of Corti in a cross section of the cochlear partition; key elements are labeled. Sound stimulates basilar membrane motion; hair cells sit in the organ of Corti (shaded) on top of the basilar membrane and are stimulated when the reticular lamina moves back and forth relative to the overlying tectorial membrane. Auditory information is relayed to the central nervous system from the inner hair cells only. Cells in the stria vascularis are responsible for the high K^+ and extracellular potential of cochlear endolymph.*

the hair bundles through the otolithic membrane, hair bundles are deflected and transduction ensues. In the semicircular canals, fluid movement within the canals, elicited during rotary head movements, displaces the cupula, which in turn displaces the hair bundles.

Key Points

- Sound energy excites cochlear hair cells by oscillating the basilar membrane.
- Basilar membrane motion peaks at a specific point along the cochlea for a given stimulus frequency.
- Oscillation of the basilar membrane causes shear between the reticular lamina and tectorial membrane.
- Hair bundles are deflected by this shear.
- Inner hair cells are probably stimulated by fluid flow arising due to basilar membrane (and tectorial membrane) motion.
- Vestibular hair cells are stimulated by net displacement of the overlying structure with respect to the hair cell.

HAIR CELL TRANSDUCTION

Bending the hair bundle initiates mechanoelectrical transduction, the conversion of a mechanical stimulus into an electrical signal that can be propagated within the nervous system. Transduction is extraordinarily sensitive, permitting detection of sound stimuli of very small magnitude.

Hair bundle bending controls the opening or closing of transduction channels, ion channels that are the central element of the hair cell's transduction apparatus. In a hair cell at rest, transduction channels spend ~10% of their time open. If a hair bundle is deflected toward the tallest stereocilia, transduction channels spend more time open; if the bundle is bent in the opposite way, toward the shortest stereocilia, transduction channels close. If the bundle is deflected perpendicular to this central axis, the opening of channels is unaffected. This directionality of transduction allows the hair cell to respond in a polarized fashion to static stimuli, which is particularly important for vestibular hair cells. Cochlear hair cells are arranged so that sound stimuli induce deflection along the sensitive axis; nevertheless, the asymmetry between the excitatory and inhibitory responses (see below) is crucial for hair cell function.

Transduction channels are very few in number, perhaps as few as one or two per stereocilium. Nevertheless, these channels are sufficiently permeable to cations that a modest change in their opening can significantly influence the membrane potential of a hair cell.

The pores of transduction channels are large in diameter, permitting the entry of most small cations. Potassium (K^+) and sodium (Na^+) ions freely enter; calcium ions (Ca^{2+}) both permeate and block this channel, reducing total current but permitting entry of this key cation. Anions do not pass through transduction channels.

Because endolymph, the extracellular fluid bathing the apical surface of a hair cell, is high in K^+ and low in Na^+, the major current-carrying cation is K^+. Because

K^+ entry typically hyperpolarizes cells, the statement that K^+ entry depolarizes hair cells often confuses students. In the hair cell, when transduction channels are open, the key electrochemical gradient is that of K^+ across the apical membrane. Although the K^+ concentration is more or less identical across this membrane, there nonetheless is a substantial electrochemical gradient because of the electrical driving force. The inside of the cell is −50 to −60 mV or so relative to ground, and the potential of the endolymphatic compartment ranges from ~0 mV (vestibular system) to +80 mV (cochlea), so the total electrical driving force can approach 150 mV. K^+ can thus effectively depolarize the hair cell.

Gating of transduction channels is extremely fast; they begin to open within a few microseconds upon application of a stimulus. This speed places severe constraints on the mechanism of gating. For instance, second-messenger systems are far too slow to accommodate microsecond-scale gating. Instead, transduction is thought to be direct: movement of the hair bundle directly tugs open the gate of the channel, without intervening biochemical steps.

The relationship between displacement and channel opening can be described by an asymmetric function, such that excitatory stimuli lead to substantially greater current entry (and hence depolarization) than the amount of current suppressed during an inhibitory stimulus. Put another way, during a symmetrical oscillating stimulus, total depolarizing current entry increases compared with that seen at rest. This is a key point, because the membrane properties of the hair cell filter the cycle-by-cycle changes in membrane potential at frequencies above a few kilohertz. Hair cells thus rely on their asymmetric displacement-channel opening relationship to ensure transfer of high-frequency auditory information.

Transduction channels are thought to be gated by tip links, fine filamentous strands that connect a short stereocilium with its neighbor. Tip links are found only along the axis of mechanical sensitivity; they are arranged such that if the bundle is deflected in the excitatory direction, they will stretch, and if the bundle is deflected in the inhibitory displacement, they slacken. Tip links thus are likely to be the elastic gating springs that control channel opening and closing (or the tip links are connected in series to the actual gating spring).

An adaptation mechanism allows hair cells to reset their sensitivity during static stimuli so that they remain responsive to small stimuli. Adaptation is thought to be more relevant for the vestibular system, which needs to filter gravitational stimuli, than the auditory system. Nevertheless, the mechanism for adaptation may be universal in hair cells because it probably functions to position transduction channels so they are optimally sensitive. In cochlear hair cells, feedback from the adaptation system may be too slow to see readily.

Key Points

- Hair bundle deflection opens transduction channels, nonselective cation channels with high conductance.
- Open transduction channels effectively depolarize the hair cell, permitting neurotransmitter release.
- The opening and closing of transduction channels is modulated by the component of the stimulus that lies parallel to the bundle's morphological axis.
- Only a few channels are found in each stereocilium.
- Channels are opened very rapidly, too rapidly for second-messenger action.
- Tip link tension probably controls the opening and closing of transduction channels.
- Adaptation ensures that tip link tension is optimal at all times.

THE COCHLEAR AMPLIFIER

One of the more astonishing achievements of the auditory system is the amplification it performs on incoming sound. Another remarkable feature is the ability of the cochlea to distinguish sounds of very similar frequency. Both of these characteristics arise from the behavior of the cochlear amplifier.

The highly viscous aqueous environment that hair bundles exist in overdampens their motion and prevents resonant behavior from arising. Simply put, water is too viscous for maximal bundle movement at the appropriate frequency. To overcome this problem, the cochlea actively compensates for this dampening by supplying energy that facilitates bundle movement. Bundle movement is augmented on a cycle-by-cycle basis to improve sensitivity, much as a correctly placed swing of the legs helps a child to swing higher and higher on a playground swing.

The amplification of basilar membrane motion by an active cochlea can be 100-fold or more, and the lower the stimulus size, the greater the amplification. Furthermore, at a given place along the basilar membrane, the cochlea only amplifies incoming stimuli at a precise frequency, the characteristic frequency of that spot. Amplification of only a narrow frequency range

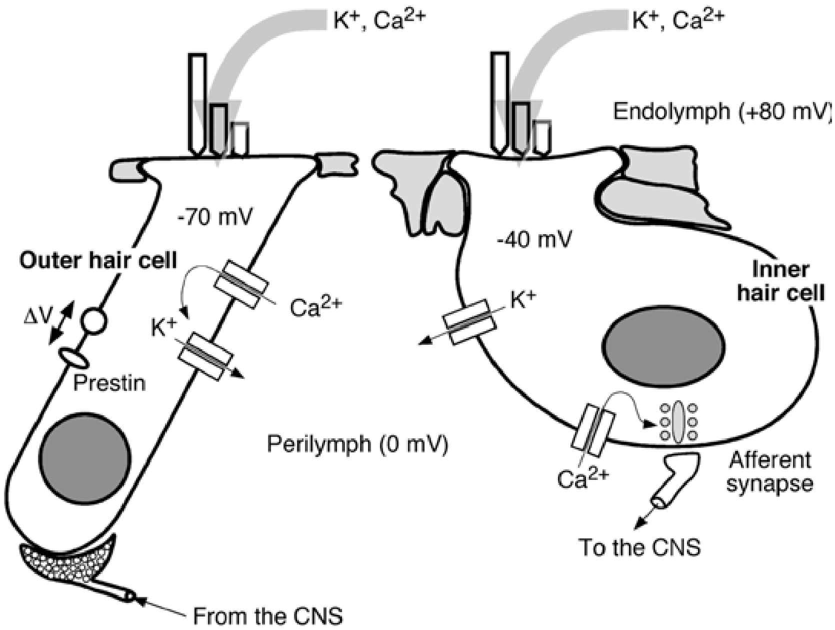


Figure 25–4 *Current and ion flow in inner and outer hair cells. K^+ enters outer hair cells via transduction channels and depolarizes the cell; ΔV depolarization in turn shortens the hair cell by affecting the outer hair cell motor (prestin). Ca^{2+}-activated K^+ channels, activated through $\alpha 9$ ACh receptors, prevent depolarization. In inner hair cells, depolarization induced by transduction increases neurotransmitter release by increasing Ca^{2+} entry at the afferent synapse.*

allows individual hair cells to specialize for a specific frequency.

The mechanism by which the cochlear amplifier functions remains uncertain. Nevertheless, it is clear that the cochlear amplifier requires functional outer hair cells. Selective destruction of outer hair cells or stimulation of efferents, which primarily innervate outer hair cells, leads to reduced sensitivity and detuning of the basilar membrane response.

The best candidate for the cochlear amplifier mechanism is active length changes of the outer hair cell body that can be elicited by changes in membrane potential (**Fig. 25–4**). Outer hair cells contract or elongate in response to depolarization or hyperpolarization, respectively, of their membranes; the length change can be as much as several percent of the total length of the cell and occurs as fast as voltage can be changed within the cell.

The lateral surface of an outer hair cell is covered with a high density of the motor protein prestin, related to an anion-transporter family, which apparently changes its surface area in response to membrane potential. Area changes of the membrane are converted into length changes of the cell because the structure of the cortical cytoskeleton restricts radial swelling of the cell and dictates axial movements.

An excitatory basilar membrane movement deflects hair bundles, depolarizing hair cells. A properly timed contraction of the cell, in response to depolarization, could coincide with the opposite phase of the stimulus and could thus augment its movement of the basilar membrane in the opposite direction.

The major problem with the outer hair cell contraction model for explaining the cochlear amplifier is that cell motility relies on membrane potential, yet the electrical filtering properties of the hair cell membrane should greatly attenuate cycle-by-cycle changes in cell potential during high-frequency stimuli. Like other cells, the hair cell's membrane can be modeled as a resistor-capacitor (RC) circuit; depending on the values of R and C, a voltage response can lag significantly behind a current step used to elicit it. In the case of the outer hair cell, voltage responses should lag behind current injection by a millisecond or more, which would effectively eliminate a substantial cycle-by-cycle membrane potential change in response to a high-frequency stimulus. Other mechanisms may compensate for this problem, or the somatic motility of the outer hair cell may serve some other purpose.

Key Points

- The cochlear amplifier can amplify small basilar membrane movements by 100-fold or more.
- At a given point along the cochlea, amplification occurs only over a very narrow frequency range.
- Outer hair cells are responsible for amplification.
- Length changes of outer hair cells that can be driven rapidly by changes in membrane voltage may account for the cochlear amplifier.
- This mechanism faces the problem, however, that cycle-by-cycle changes in membrane potential arising from transduction currents will be greatly diminished at high frequencies.

SYNAPTIC TRANSMISSION

Hair cell excitation is completed with afferent signaling to cranial nerve (CN) VIII. In addition, hair cell function is modulated by efferent signaling, which hyperpolarizes hair cells and may reduce the gain of the cochlear amplifier.

Depolarization of the hair cell membrane during excitatory mechanoelectrical transduction leads to the opening of voltage-sensitive Ca^{2+} channels, which are located along the basal (auditory) or basolateral (vestibular) surface of a hair cell, at the synaptic release sites. These channels are L-type, dihydropyridine-sensitive channels that do not undergo Ca^{2+}-dependent inactivation; this lack of adaptation may be an accommodation to the hair cell, which is tonically active, constantly releasing neurotransmitter. Entering Ca^{2+} triggers neurotransmitter release; this transmitter, which is either glutamate or a chemically related compound, diffuses to excitatory postsynaptic sites and initiates an action potential in the VIIIth CN fiber.

Synaptic release sites have a distinctive morphology that is reminiscent of the ribbon synapses of photoreceptors. In hair cells, a halo of small, clear synaptic vesicles surrounds an electron-dense sphere of unknown function. Hair cells are capable of releasing transmitter at a high rate for prolonged periods; this synaptic sphere may play a role in the high release rate.

Efferent synapses also contact hair cells. In the cochlea, the great majority of the efferents directly contact outer hair cells, where they appear to modulate the gain of the cochlear amplifier.

Direct evidence from lower vertebrates and indirect evidence from mammals suggest that efferent stimulation leads to hyperpolarization of outer hair cells. Because hyperpolarization in this case is associated with an increased K^+ conductance along the basolateral surface, the depolarizing effect of opening transduction channels is reduced. That is, the receptor potential arising from a given transduction current is substantially diminished by efferent activation. Because the gain of the cochlear amplifier may depend on the magnitude of the outer hair cell receptor potential, the sensitivity of the cochlea is reduced.

The sensitivity of inner hair cells is also reduced by efferent stimulation, but that appears to be an indirect effect arising from the decline in gain of the cochlear amplifier.

The efferent neurotransmitter is acetylcholine (ACh); it appears to bind to a rare receptor type, the $\alpha 9$ ACh receptor, which is a cation-conducting receptor channel. In lower vertebrates, hyperpolarization ensues when this receptor is activated because Ca^{2+} entering through this receptor channel activates a class of Ca^{2+}-activated K^+ channels, of much larger total conductance than the $\alpha 9$ channels. Open K^+ channels hyperpolarize the cell, moving the membrane potential to the K^+ equilibrium potential.

Key Points

- Depolarization opens voltage-sensitive Ca^{2+} channels.
- Entering Ca^{2+} triggers neurotransmitter release.
- Afferent synapses are optimized for continuous and rapid release of neurotransmitter.
- Efferent stimulation hyperpolarizes hair cells, reducing the transmission of sound information through the auditory system.

Suggested Readings

Geisler CD. From Sound to Synapse: Physiology of the Mammalian Ear. New York: Oxford University Press; 1998

Gillespie PG, Walker RG. Molecular basis of mechanosensory transduction. Nature 2001;413:194–202

Hudspeth AJ. How the ear's works work. Nature 1989;341:397–404

Pickles JO, Corey DP. Mechanoelectrical transduction by hair cells. Trends Neurosci 1992;15:254–259

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Hair cells resemble
   A. Neurons
   B. Epithelial cells
   C. Both of the above
   D. Neither of the above

2. Hair cell transduction ensues because
   A. Adjacent stereocilia slide along each other, stretching basal linkages immediately above the cuticular plate.
   B. Transduction channels open, letting in Na^+ to depolarize the cell.
   C. Stereocilia bend along their lengths, opening membrane channels that are anchored in the stereociliary actin filaments.

D. Hair bundles are deflected back and forth because of an up-and-down movement of the cuticular plate.
E. None of the above

3. The high concentration of K^+ and large extracellular potential of endolymph means
   A. Hair cells hyperpolarize when K^+ enters transduction channels.
   B. Cells other than hair cells are responsible for the primary metabolic loads placed on the organism by transduction.
   C. Transduction channels open much faster than they would with an Na^+-rich environment.
   D. Transduction channels open much faster than they would if the endocochlear potential was zero.
   E. None of the above

4. Afferent neurotransmitter release
   A. Uses γ-aminobutyric acid (GABA) as the neurotransmitter
   B. Is inhibited by hair cell depolarization
   C. Is not readily fatigued
   D. Employs the $\alpha9$ acetylcholine (ACh) receptor
   E. None of the above

Chapter 26

AUDITORY PROCESSING IN SENSORINEURAL HEARING LOSS

M. Charles Liberman

From a physiological perspective, a great deal is known about the way in which environmental sounds, including speech, are "coded" by neurons in the auditory pathway. There is also a significant amount known about the ways in which damage to the peripheral auditory system, particularly to the inner ear, changes this normal pattern of neural activity. The purpose of this chapter is to show, in simple terms, how such understanding of the structural and functional changes in the auditory periphery can help explain several puzzling aspects of sensorineural hearing impairment. Those aspects explicitly dealt with will include (1) the phenomenon of loudness recruitment in which the perception of loudness grows with exceptional rapidity compared with normal, (2) the observation that individuals with identical pure-tone audiograms can have dramatically different auditory capabilities, (3) the fact that sensorineural hearing loss can be difficult to rectify with amplification provided by a hearing aid, and (4) the common presence of tinnitus, or ringing in the ears, in cases of sensorineural hearing loss.

NORMAL STRUCTURE AND FUNCTION OF THE EAR

Hair Cells: Transduction and Synaptic Transmission

Sounds begin as air pressure waves, which vibrate the eardrum. These vibrations are conducted to the spiraling inner ear by a series of three bony ossicles, the innermost of which (stapes) moves in and out of the inner ear like a piston. These pumping movements generate pressure waves in the fluids filling the inner ear, or cochlea. The cochlea, in turn, transduces these pressure waves into electrical activity in the auditory nerve,

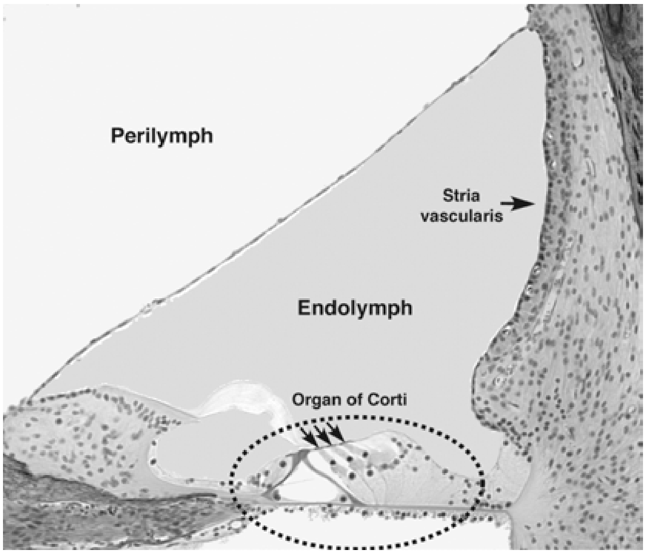


Figure 26–1 *A light micrograph of a cross section through the cochlear duct. Scale bar = 100 μm.* (See **Color Plate 26–1**.)

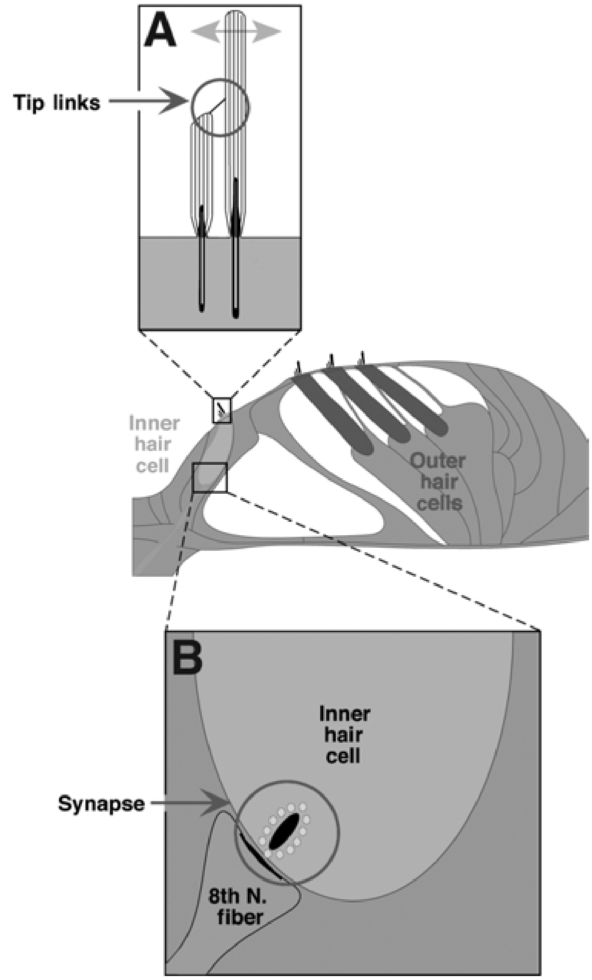


Figure 26–2 *Schematic illustration of the structural features underlying **(A)** transduction and **(B)** synaptic transmission within the organ of Corti.* (See **Color Plate 26–2**.)

which relays the information to the brain. For a review, see Geisler (1998).

The transduction process in the cochlea requires cooperative function of several different cell types lining the cochlear duct (**Fig. 26–1**). This duct is filled with endolymph, an extracellular fluid rich in potassium (K^+) and low in sodium (Na^+) (more like typical intracellular fluids). The endolymphatic space also maintains a large electric potential of +100 mV with respect to the surrounding perilymphatic spaces. Both the unique ionic composition and the electric potential of the endolymphatic space are maintained by a group of cells known as the stria vascularis.

In humans, the cochlear duct spirals for roughly 35 mm from the base of the cochlea (near the stapes) to the apex (Schuknecht, 1974). The size, mass, and stiffness of many of its cellular elements, especially those in the organ of Corti, change systematically from one end of the spiral to the other. This yields a mechanical "tuning" such that pressure waves produced by high-frequency sounds cause the organ to vibrate at the base of the spiral (near the stapes), while low-frequency sounds cause vibrations at the apical end. Thus, a given sound causes vibrations of only those regions of the organ of Corti that are tuned to the frequency components of the sound, and the inner ear acts as a mechanical frequency analyzer.

The transduction process, per se, is performed by two classes of sensory cells within the organ of Corti, the inner hair cells (IHCs) and the outer hair cells (OHCs). Sound-induced pressure waves in the cochlear fluids bend the sensory "hairs," called stereocilia, on top of the hair cells.* As schematized in **Fig. 26–2**, this bending alternately stretches and slackens the tip links that join adjacent stereocilia. When the tip link stretches, it directly opens ion channels in the stereocilia membrane and allows K^+ to enter the hair cell from the endolymph, driven by its considerable electrical gradient (+100 mV in the endolymph to −60 mV within the hair cell). This influx of positively charged K^+ ions changes the electrical potential within the hair cell, which causes release of chemical neurotransmitters from synaptic vesicles at the base of the hair cell (**Fig. 26–2**). Auditory nerve fibers, contacting the hair cells, respond to the neurotransmitter by producing action potentials, spikes of electrical current, which propagate along the nerve fibers to reach the brain within a few thousandths of a second. Patterns of electrical activity across the

*These stereocilia are neither hairs nor cilia; rather, they are modified microvilli.

40,000 fibers of the auditory nerve are "decoded" by the brain, resulting in the sense we call hearing.

IHCs and OHCs play fundamentally different roles in inner ear function. Almost all auditory nerve fibers contact only IHCs (Spoendlin, 1969). The IHCs are simple transducers, changing mechanical energy into electrical energy (i.e., pressure waves into action potentials). The OHCs are tiny amplifiers that can enhance the mechanical vibrations of the organ of Corti (Brownell, 1983). This OHC contribution is crucial to the normal sensitivity and frequency selectivity of the inner ear (Dallos and Harris, 1978). As we will see, the loss of OHCs underlies many of the most common problems in sensorineural hearing loss.

Auditory Nerve Response: Rate, Synchrony, and Frequency Tuning

In the human cochlea, there are ~3500 IHCs, each of which is contacted by ~10 auditory nerve fibers (Nadol, 1983). When a single auditory nerve fiber (ANF) is impaled with a microelectrode, one finds that, like most other neurons, ANFs communicate with each other via trains of action potentials (Kiang et al, 1965). As schematized in **Fig. 26–3**, ANFs can carry information about acoustic events to the brain in one of two ways: by changing their spike rate or by synchronizing their spikes to individual peaks in a stimulus waveform. For a review of ANF response properties, see Geisler (1998).

Consider first the changes in spike rate. In both parts A and B of **Fig. 26–3**, a 50 msec tone burst is schematized (in blue), beginning after a 50 msec silent period. Sample spike trains of ANFs are shown in red. In both cases, the neuron discharges, even during the silent periods. This background or resting discharge is called spontaneous activity. In the case shown in **Fig. 26–3A**, the spontaneous rate is two spikes per 50 msec, or 40 spikes per sec. In fact, spontaneous activity in ANFs can be higher than 100 spikes per sec. In both parts A and B, the neuron is "responding" to the tone burst with an increase in average rate: for example, in part A, the spontaneous rate of 40 spikes per sec rises to an average sound-driven rate of 140 spikes per sec (seven spikes in 50 msec) during the time when the tone is on.

Synchronization in auditory nerve discharge is illustrated in **Fig. 26–3A**. Synchrony, in this context, means that the spike discharges tend to be time-locked with individual cycles of the stimulus (in this case, a pure tone). Thus, in **Fig. 26–3A**, spikes 3, 4, 6, 7, 8, and 9 all occur immediately after a positive zero-crossing in the stimulus waveform. Note that not all spikes are synchronized (e.g., spike 5), and there is not necessarily a spike for every cycle of the stimulus. Nonetheless, if all the intervals between spikes were measured, there would be a predominance of intervals corresponding to the stimulus period (and integer multiples thereof). Such an interval analysis could (with appropriate central processing circuitry) directly reveal information as to the

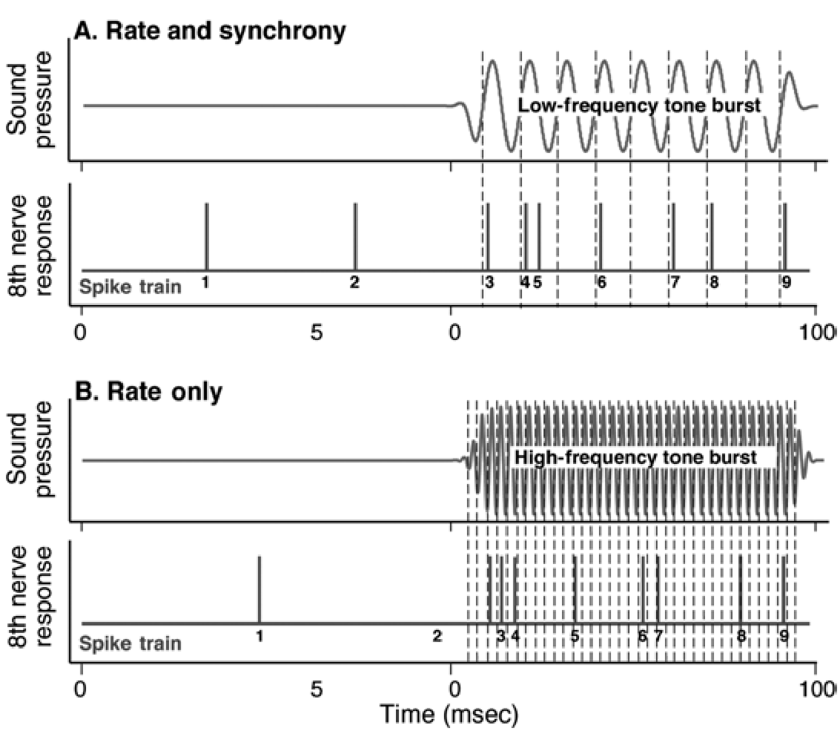


Figure 26–3 (A,B) *Schematic illustration of the two major ways in which acoustic information is coded in auditory nerve response: changes in average discharge rate and changes in degree of synchronization.* (See **Color Plate 26–3**.)

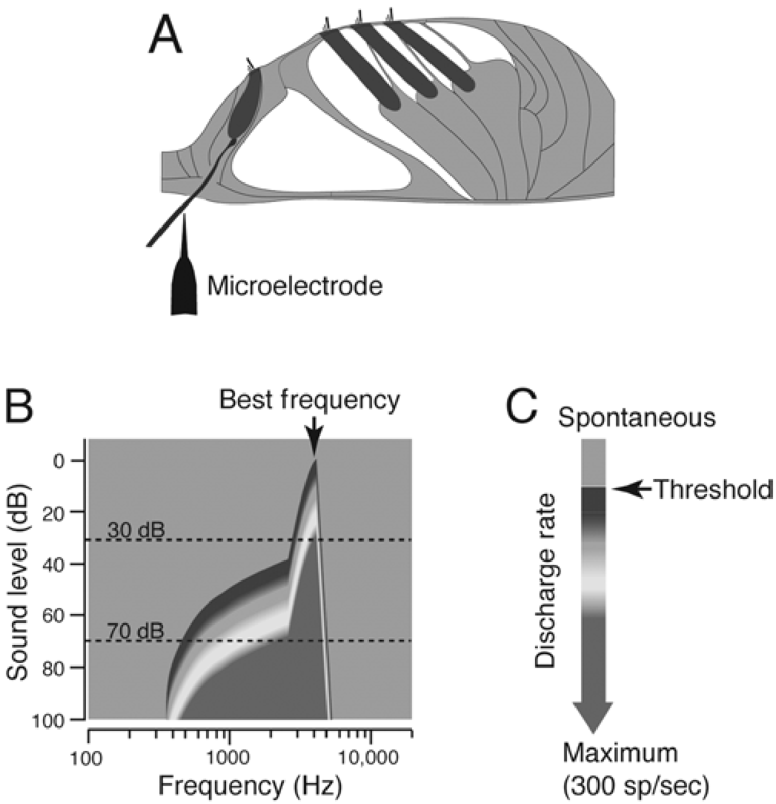


Figure 26–4 ***(A)*** *Schematic representation of the response area of a single auditory nerve fiber indicating* ***(B)*** *extreme frequency selectivity at low sound pressure levels.* ***(C)*** *Rate of discharge in this fiber is indicated via a pseudo-color code.* (See **Color Plate 26–4**.)

frequency of the stimulus. Synchrony arises only at low frequencies (<4 kHz) because only at low frequencies can all the electrical events involved in transduction and synaptic transmission faithfully follow the individual cycles of a sinusoid (Johnson, 1980). To simplify matters, in the discussion that follows, we will consider only a relatively high-frequency stimulus (4 kHz) and thus will restrict our attention to rate changes in auditory nerve fibers.

Individual ANFs show frequency selectivity in their response (Kiang et al, 1965), reflecting the mechanical tuning of the organ of Corti already described (Ruggero and Rich, 1991). The range of frequency/level combinations of sound to which a fiber responds is called its response area; the response area of one fiber is schematized in **Fig. 26–4B**. This response-area graph has axes like an audiogram: sound frequencies, in Hertz (Hz),† are arrayed along the *x*-axis from low to high, and sound intensities, in decibels (dB),‡ are plotted on the *y*-axis with lower sound levels toward the top. In this figure, the strength of the fiber's response is coded in pseudo-color (see **26–4C**): red for maximum excitation (~300 spikes per sec), with yellows, greens, and blues indicating a steadily weaker response. Spontaneous activity is shown as gray. At low sound levels (e.g., 0–30 dB), only a narrow range of tone frequencies excites this fiber. At higher sound levels (e.g., 70 dB), a wider range of frequencies is excitatory, with the activation spreading more to low- than to high-tone frequencies. The tone frequency to which a fiber is most sensitive, its best frequency, pinpoints where along the spiraling organ of Corti that fiber contacts a hair cell: high best frequencies characterize neurons from the cochlear base, low best frequencies those in the apex.

Dynamic Range and Loudness Coding

A key point of the pseudo-color representation in **Fig. 26–4** is the limited "dynamic range" of a single ANF. As illustrated, a fiber attains its maximum discharge rate (the red portion of the response area) within only ~30 dB above threshold at any particular frequency. Thus, for any one fiber, there is a limited range of sound

†Hertz (Hz), which is the same as cycles per second. For point of reference, middle C on the piano is 240 Hz. The human ear hears sounds from frequencies as low as about 50 Hz to frequencies as high as about 25,000 Hz.

‡Stimulus level is expressed on a logarithmic scale, the decibel (dB) scale, because the ear operates over a very wide range of sound amplitudes. For every 20 dB increase in sound pressure, the amplitude of the sound waves increases by a factor of 10. Thus, between threshold levels (~0 dB) and the threshold of auditory pain (~140 dB), there is a 10^7-fold range of sound amplitudes.

pressures over which response rate can vary, and thereby code information. Put another way, the response area in **Fig. 26–4** shows a wide range of tone frequencies and sound levels that will produce the same degree of maximal excitation in a single ANF.

How, then, can the typical human listener perceive loudness over a dynamic range of ~100 dB if the dynamic range of single ANFs is only 30 dB? One view is suggested by the schematic in **Fig. 26–5**: by looking simultaneously over the entire array of ANFs from base to apex, information can be obtained about stimulus level over a wide range of intensities by analyzing the spread of excitation along the cochlear spiral. To illustrate this point graphically (**Fig. 26–5**), we consider five ANFs, chosen to be equally distributed along the cochlear duct from apex to base. These five regions are schematized by five cross sections through the organ of Corti, spaced along an uncoiled cochlear spiral and steadily decreasing in size from apex to base. In the top panel, we schematize their respective response areas, with best frequencies almost spanning the audible frequency range from 200 Hz to 30 kHz.

In this simple example, we consider a 4 kHz tone presented at three different sound pressure levels: 30, 60, and 90 dB sound pressure levels (SPLs). Consider first the 30 dB tone. As can be seen in the top panel, the 30 dB tone is only within the response area of one fiber: the one tuned to 4 kHz. Thus, in the uncoiled cochlea labeled 30 dB, we show "response" (nongray arrow) in only the middle fiber. For that fiber the response intensity is orange, reflecting a moderate level of excitation. As the stimulus level rises to 60 dB, the firing rate in this fiber increases to maximum. In addition, activity begins to be seen in a more basal fiber. Similarly, as intensity rises to 90 dB, the activity spreads still farther toward the base. Clearly, important information about stimulus intensity is available by assessing the spread of excitation along the cochlear spiral.

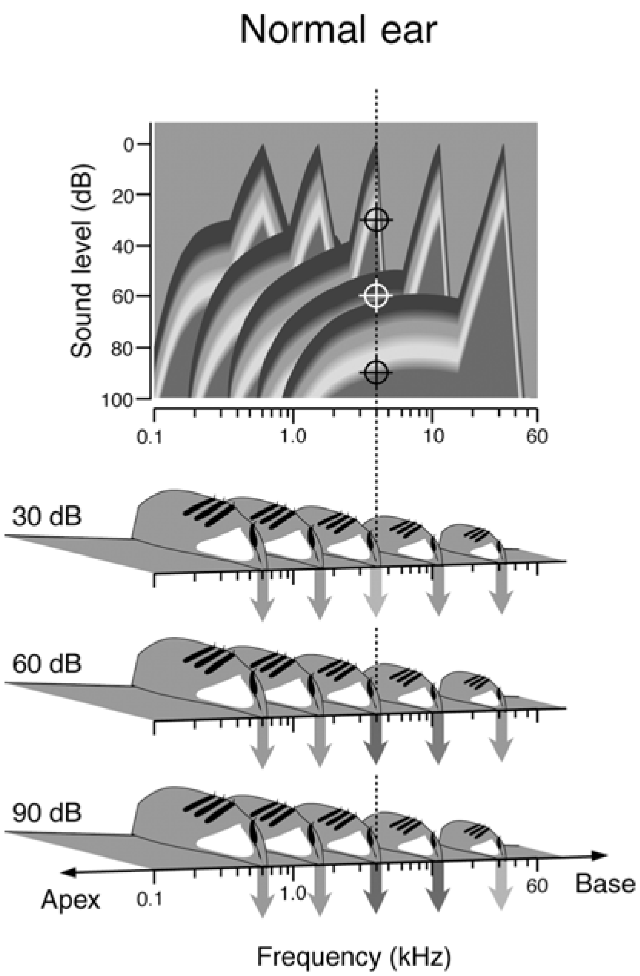


Figure 26–5 *Schematic representation of the response areas of five nerve fibers in a normal cochlea (top) and the discharge rates seen across this fiber ensemble to a 4 kHz tone presented at either 30, 60, or 90 dB sound pressure level (below). The pseudo-color code for rate is the same as that shown in **Fig. 26–4C**.* (See **Color Plate 26–5**.)

EFFECTS OF INNER EAR DAMAGE ON AUDITORY FUNCTION

Most structures of the inner ear are susceptible to damage; however, the sensory cells are the most vulnerable elements, particularly the OHCs (Liberman and Kiang, 1978). For a wide variety of cochlear insults ranging from overexposure to loud sounds to treatments with ototoxic drugs, the predominant structural change is a loss of OHCs. Correspondingly, the great majority of individuals with sensorineural hearing loss have significant loss of, or damage to, the OHCs.

Fig. 26–6B schematizes the response area of a nerve fiber located in a region of OHC loss. Without the OHCs to amplify the organ of Corti's vibrations, much higher sound levels are required to excite the fiber, compared with the normal response in part A. The sharply tuned, sensitive portion of the normal fiber's response area is gone, resulting in ~40 dB of hearing loss at the fiber's best frequency, and greatly decreased frequency selectivity. Note, however, that spontaneous activity (gray) and maximum discharge rate (schematized as red) are unchanged from normal. Thus, at high sound levels, the response area is normal in all respects (Kiang et al, 1970).

In more severe cases of sensorineural hearing loss, there can be IHC loss in addition to OHC loss (loss of IHCs only is virtually unknown). In a region of complete hair cell loss, as schematized in **Fig. 26–6C**, the nerve terminals degenerate within the organ of Corti (Johnsson, 1974). These fibers will be silent: they will not respond to

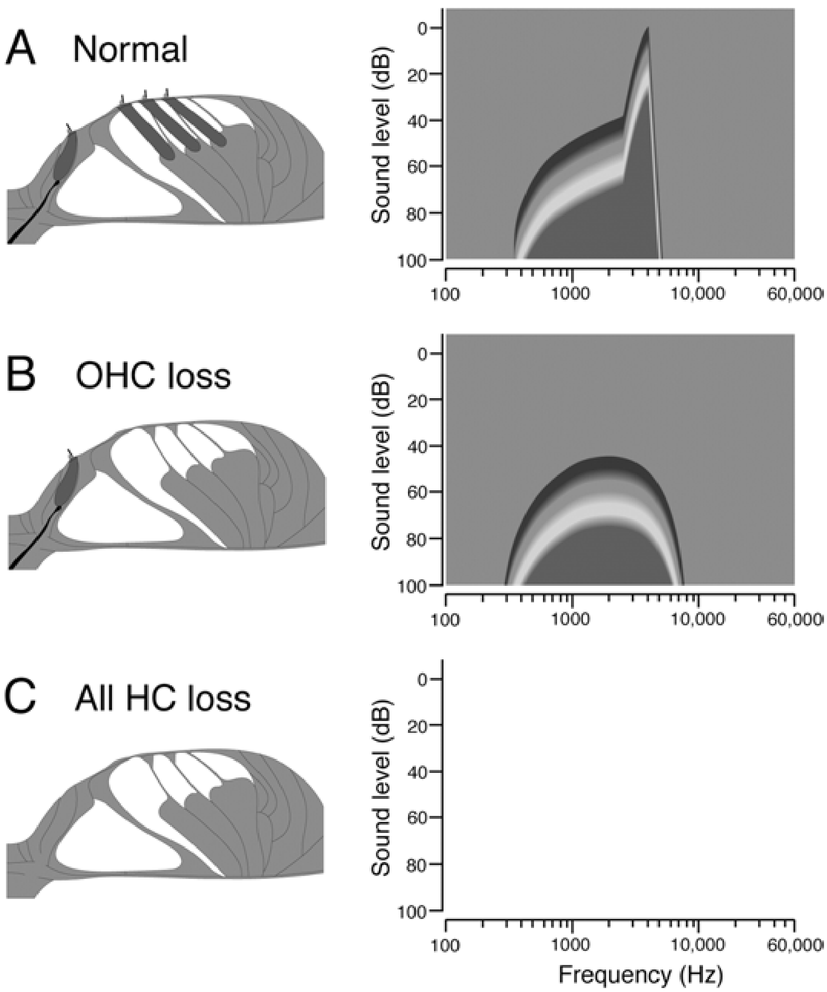


Figure 26–6 *Schematic representations of the response areas of three different auditory nerve fibers:* ***(A)*** *from a normal ear,* ***(B)*** *from a region with selective outer hair cell (OHC) loss, and* ***(C)*** *from a region with loss of both inner hair cells (IHCs) and OHCs.* (See **Color Plate 26–6**.)

sound and will have no spontaneous activity (Liberman and Kiang, 1978). Thus we schematize their response areas as a blank white. Of course, so long as the cell body and central axon remain, the neuron will likely remain excitable by electrical stimulation such as that provided by a cochlear prosthesis.

Outer Hair Cell Loss and Loudness Recruitment

To understand how inner ear damage disrupts normal sound perception, we need to consider the pattern of sound-evoked activity across the entire spectrum of fibers from base to apex. **Fig. 26–7** schematizes the effect of an OHC lesion restricted to the midfrequency region of the cochlea. Such a lesion might be caused by a mild acoustic trauma, which classically causes a midfrequency hearing loss (Schuknecht, 1974). As for the normal ear schematized in **Fig. 26–5**, **Fig. 26–7** illustrates the response areas of five selected neurons, including one from the midfrequency region in which the OHC loss has occurred (red circle in **Fig. 26–7C**). The response area of that neuron shows the loss of its sensitive tip, as already discussed; the other four neurons show normal response areas. Superimposed on the response areas in **Fig. 26–7A** are a series of five points indicating the lowest sound pressure at each of these frequencies for which any neural activity would be expected. On the assumption that the perception of a tone requires activity in only a small population of ANFs, we argue that this envelope of the response areas should correspond very closely to the behavioral audiogram (as shown in **Fig. 26–7B**). Thus we expect that the midfrequency loss of OHCs should result in a midfrequency notch in the audiogram.

In **Fig. 26–8**, we illustrate schematically why OHC loss might be expected to be associated with loudness recruitment. Once again, we consider the response in five selected auditory neurons to a 4 kHz tone at 30, 60, or 90 dB SPL. On the right side, the OHC loss truncates the response area of the midfrequency fiber, so that the low-level tone (30 dB) now elicits no response. Thus the five-neuron array on the damaged side shows only spontaneous activity (gray arrows). This is the basis for the hearing loss as seen on an audiogram. When the level is

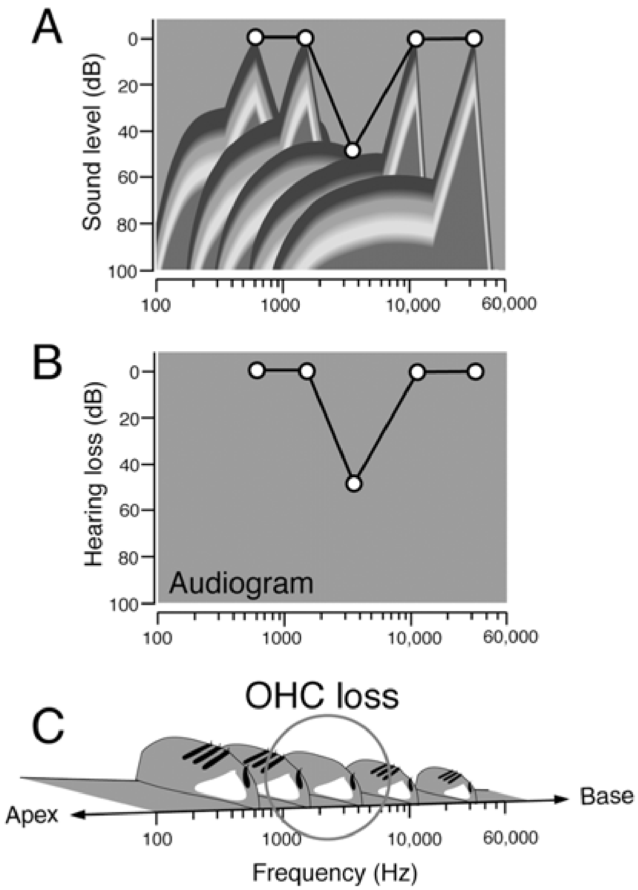


Figure 26–7 ***(A)*** *Schematic representation of the response areas of five fibers in* ***(C)*** *an ear with a restricted OHC lesion in the middle of the cochlea.* ***(B)*** *The audiogram associated with such a lesion pattern.* (See **Color Plate 26–7**.)

raised to 60 dB, the sound is now within the response area of two fibers, as in the normal ear; however, the response rate is lower than normal in the damaged neuron (green rather than red). Finally, when sound pressure is increased to 90 dB, even the damaged fiber now discharges at its maximum rate, producing an overall activity pattern identical to that in the normal ear. If the patterns of activity are normal, then the auditory perception, including the perceived loudness, should be normal. This provides a simple explanation for loudness recruitment, where a damaged ear can show a threshold shift yet maintain a normal loudness sensation at high sound pressures. The key factor is that loss of OHCs increases thresholds and broadens tuning without destroying ANFs or changing their maximum discharge rate or the limited dynamic range of individual neurons.

Inner Hair Cell Loss, Tinnitus, and Pitch Shifts

Let us now add the effects of IHC loss, and consider neuronal activity in an ear with a region of total hair cell loss, as schematized in **Fig. 26–9**. Such a lesion, indicated by the circled region **Fig. 26–9C**, might also be expected to arise from acoustic trauma, perhaps a more severe exposure than that which caused the selective OHC loss illustrated in **Fig. 26–7**. As before, we infer the audiogram by estimating the lowest sound pressure at each test frequency that might be expected to evoke spike activity in any ANFs. Note that the midfrequency notch in this inferred audiogram is very similar to that

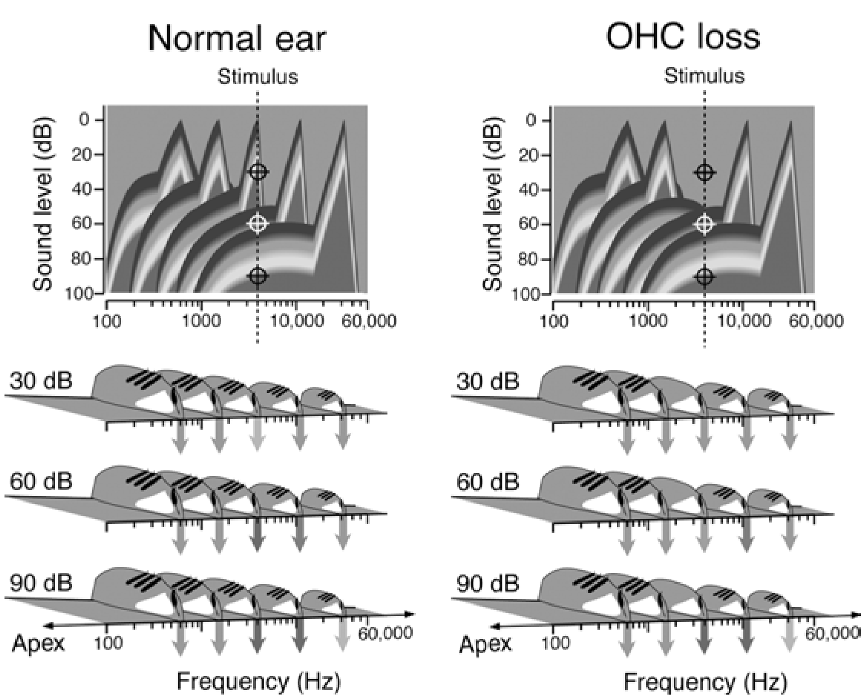


Figure 26–8 Comparison of the ensemble discharge patterns across five representative auditory nerve fibers for a normal ear versus an ear with the type of midcochlear OHC loss shown in ***Fig. 26–7****. All other display conventions of the figure are as described for* ***Fig. 26–5****.* (See **Color Plate 26–8**.)

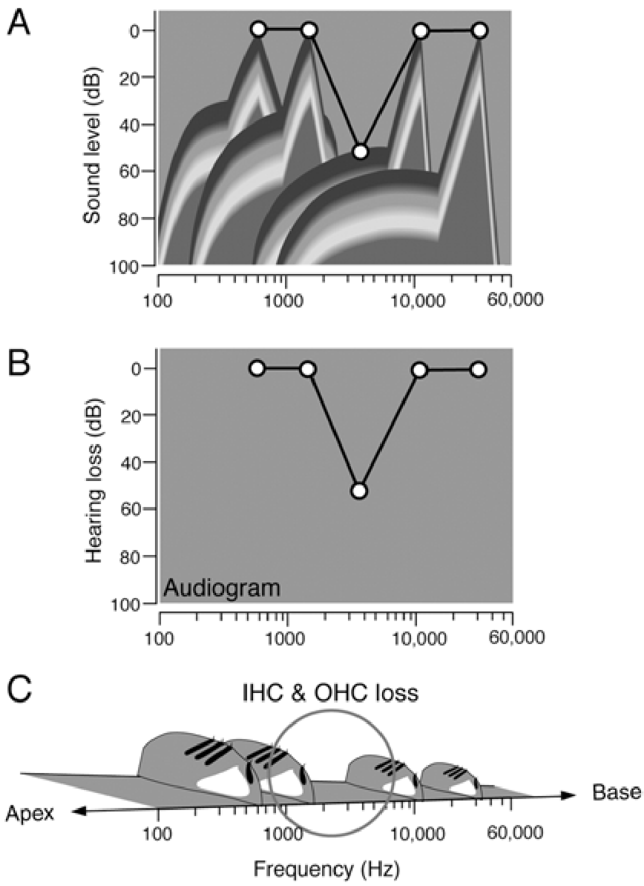


Figure 26–9 ***(A)*** *Schematic representation of the response areas of five fibers in* ***(C)*** *an ear with a restricted combined IHC/OHC lesion in the middle of the cochlea.* ***(B)*** *The audiogram associated with such a lesion pattern.* (See **Color Plate 26–9**.)

seen in the case of selective OHC loss (**Fig. 26–7**). Although the midcochlear region is completely destroyed, spread of excitation, as the 4 kHz tone intensity is increased, ensures that more basally located fibers begin to be activated, and the ear shows moderate threshold elevation rather than complete lack of response.

As schematized in **Fig. 26–10**, we do not expect that the ear with significant IHC loss will show complete loudness recruitment. As the pseudo-color code shows, there is no sound level of the 4 kHz tone that can evoke a normal pattern of activity across the auditory nerve array, because even at high sound pressure levels a band of fibers remains nonresponsive. Even at the highest sound pressures, the damaged ear's responses are generated by cochlear regions basal to the location actually tuned to the midfrequency stimulus.

If we assume that the perception of pitch (at least high-frequency pitch) is based on analysis of the cochlear location of ANF activity, this type of IHC loss might be expected to alter pitch perception. Indeed, pitch-matching anomalies have been documented in cases of noise-induced hearing loss. In the 1940s, several auditory physiologists exposed themselves unilaterally to very high level tones and then performed loudness and pitch matching experiments between the two ears (Davis et al, 1950).[§] In some of the subjects, the following postexposure pitch-matching anomaly was noted: when presenting a range of mid- to high-frequency tones to the damaged ear (in the frequency region of maximum threshold shift), the perceived pitch did not vary and was always matched to a high-frequency tone in the good ear. If we assume that the two ears of such a subject looked like the normal and damaged ears in **Fig. 26–10**, we can understand why any tone in the damaged ear between 4 and 10 kHz would be perceived as a 10 kHz tone: it is only the 10 kHz population of ANFs that is stimulated regardless of the input frequency.

The loss of IHCs renders the nerve fibers in that cochlear region totally nonresponsive, lacking even spontaneous activity. Thus, even in quiet, the activity pattern across the spectrum of auditory nerve fibers differs from normal: there is a part of the nerve spectrum that is silent. It is possible that such discrete zones of neural inactivity lead to the perception of tinnitus, as is commonly seen in some kinds of sensorineural hearing loss. In the normal ear, there are never adjacent regions of contrasting ANF activity levels except when a tone is present; the artificial contrast set up by IHC loss may "force" the brain to assume that a tone is present, when none is actually there. To put the hypothesis in a more biologically correct form, one could suggest that, given the prevalence of inhibitory interneurons in neural circuits within the auditory central nervous system, the loss of activity in some restricted cochlear region might well be translated into increased activity (loss of inhibition via inhibitory interneurons) somewhere in the ascending auditory pathway. Such an increased activation level could be interpreted as a sound percept, even when sounds are not present.

SUMMARY

The analysis of ANF activity patterns in normal and damaged ears presented here is greatly oversimplified, largely ignoring consideration of neuronal synchrony, which is doubtless very important to pitch perception in both normal and abnormal ears for low- and midfrequency stimuli. Nonetheless, the examples presented illustrate that the anomalies in ANF activation patterns that result from cochlear pathology predict (1) that

[§]They thought they were studying reversible hearing loss; in fact, most of the participants ended up with permanent noise-induced hearing loss.

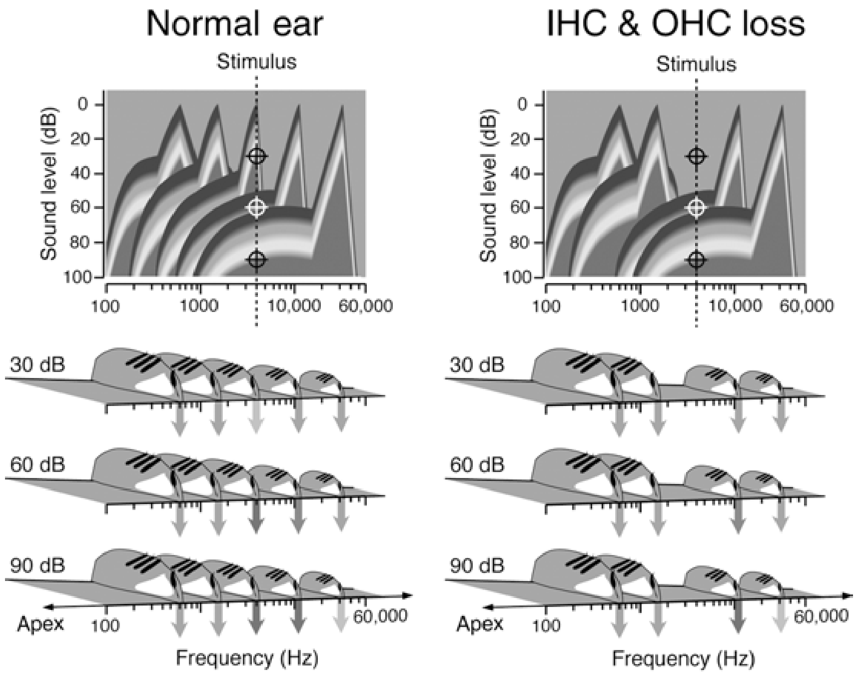


Figure 26–10 *Comparison of the ensemble discharge patterns across five representative auditory nerve fibers for a normal ear versus an ear with the type of midcochlear outer hair cell (OHC) loss shown in **Fig. 26–7**. All other display conventions of the figure are as described for **Fig. 26–5**.* (See **Color Plate 26–10**.)

selective OHC loss should give rise to recruiting types of hearing loss, (2) that IHC loss might give rise to anomalies of pitch and tinnitus, and (3) that the pure-tone audiogram can be identical in cases in which the underlying pathophysiology is very different.

The simplified analysis presented here also gives some insight into why neither OHC nor IHC loss can be compensated by simply amplifying the input signal. The brain must decode the nature of the external sound based only on the time patterns and locations of the neural activity reaching it. The effects of inner and OHC loss on these patterns are much more complex than the effects of changing sound transmission through the external and middle ears, which merely shifts the response areas of the inner ear's nerve fibers without changing their shapes. A simple shift along the sound level axis is well compensated through conventional hearing aid applications. Sensorineural hearing loss clearly presents more of a challenge.

ACKNOWLEDGMENTS

The skillful assistance of L. D. Liberman in the design and execution of the illustrations is gratefully acknowledged. Research has been supported by grants from the National Institute for Deafness and Other Communicative Disorders (NIDCD).

Suggested Readings

Brownell WE. Observations on a motile response in isolated outer hair cells. In: Webster WR, Aitkin LM, eds. Mechanisms of Hearing. Clayton, Victoria, Australia: Monash University Press; 1983:5–10

Dallos P, Harris D. Properties of auditory nerve responses in absence of outer hair cells. J Neurophysiol 1978;41:365–383

Davis H, Morgan CT, Hawkins JE Jr, Galambos R, Smith FW. Temporary deafness following exposure to loud tones and noise. Acta Oto-Laryngologica Suppl 1950;88:5–57

Geisler, CD. From Sound to Synapse. New York: Oxford University Press;1998

Johnson DH. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J Acoust Soc Am 1980;68:1115–1122

Johnsson L-G. Sequence of degeneration of Corti's organ and its first-order neurons. Ann Otol Rhinol Laryngol 1974; 83:294

Kiang NYS, Moxon EC, Levine RA. Auditory-nerve activity in cats with normal and abnormal cochleas. In: Ciba Foundation Symposium on Sensorineural Hearing Loss. London: Churchill; 1970

Kiang NYS, Watanabe T, Thomas EC, Clark LF. Discharge Patterns of Single Fibers in the Cat's Auditory Nerve (MIT Research Monograph No. 35). Cambridge: MIT Press; 1965

Liberman MC, Kiang NYS. Acoustic trauma in cats, cochlear pathology and auditory-nerve activity. Acta Otolaryngol 1978;358:5–63

Nadol JB Jr. Serial section reconstruction of the neural poles of hair cells in the human organ of corti, I: Inner hair cells. Laryngoscope 1983;93:599–614

Ruggero MA, Rich NC. Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane. J Neurosci 1991;11:1057–1067

Schuknecht HF. Pathology of the Ear. Cambridge, MA: Harvard University Press; 1974

Spoendlin HH. Innervation of the organ of Corti of the cat. Acta Otolaryngologica (Stockh) 1969;67:239–254

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. If the only structural pathology in a particular cochlear region is the total loss of outer hair cells, the type I auditory nerve fibers from that region will
   A. Degenerate
   B. Not degenerate, but will not respond to sound and will have no spontaneous activity
   C. Have normal thresholds to sound but no spontaneous activity
   D. Have thresholds elevated by 40 dB but normal patterns of spontaneous activity
   E. Have thresholds elevated by 40 dB and no spontaneous activity
2. In a noise-damaged ear, if there is complete loss of inner and outer hair cells from the 3 kHz to the 6 kHz region of the cochlea (and all other cochlear regions are completely normal), the audiometric pattern should be
   A. A 40 dB loss at 4 kHz with normal hearing at 2 and 8 kHz
   B. An 80 dB loss at 4 kHz with normal hearing at 2 and 8 kHz
   C. A profound loss at 4 kHz with normal hearing at 2 and 8 kHz
3. As the sound pressure level of pure tone is increased, the number of auditory nerve fibers responding to the tone increases primarily because
   A. Fibers with progressively higher best frequencies begin to respond.
   B. Fibers with progressively lower best frequencies begin to respond.
   C. Fibers with similar best frequencies but higher thresholds begin to respond.

Chapter 27

PATHWAYS OF HEARING AND BALANCE

ALAN D. LEGATT

The neural signals carrying auditory and vestibular information from the inner ear follow an extremely complex series of neural pathways within the brain, with axons both synapsing in and bypassing several of the relay nuclei, and partial decussations at multiple levels. The pattern of these pathways and their synaptic connections determines the effect of brain lesions on hearing, on vestibular function, and on diagnostic tests such as brainstem auditory evoked potentials (BAEPs) and caloric nystagmus that are used to localize these lesions. Adjaceat to the afferent pathways are a network of efferent or descending pathways that provide feedback and rostral control of the sensory input. Lesions of the efferent pathways also can produce clinical symptoms. This chapter describes the anatomy and function of the ascending and descending auditory and vestibular pathways and the clinical relevance of their anatomy and physiology.

Our knowledge of the cytoarchitecture and connectivity of the brainstem auditory and vestibular systems is predominantly based on studies in nonprimate experimental animals, especially cats. A smaller number of studies in primates suggest that the overall pattern is similar, though there may be some evolutionary differences, which will be mentioned later. Studies in humans have played a larger role in the characterization of cerebral cortical areas subserving auditory and vestibular function, especially language functions.

AFFERENT AUDITORY PATHWAYS

COCHLEAR NERVE

The output of the cochlea is carried by the cochlear or auditory nerve, which is part of cranial nerve (CN) VIII. The cell bodies of the cochlear nerve fibers, comprising the spiral ganglion, lie within the modiolus of the cochlea. These are true bipolar neurons with processes arising on opposite ends of the cell bodies. The short distal processes receive synapses from the cochlear hair cells. The much longer proximal processes join together to form the cochlear nerve and synapse on neurons of the cochlear nucleus, where all of the axons in the cochlear nerve terminate. Cochlear nerve axons are ensheathed by central nervous system (CNS)-type myelin, produced by oligodendrogliocytes, along most of CN VIII; the transition to peripheral nervous system–type myelin, produced by Schwann cells, occurs close to the cochlea. CN VIII is therefore vulnerable to diseases that affect CNS myelin, such as multiple sclerosis.

The cochlear nerve travels within the petrous temporal bone from the cochlea to the posterior fossa. Accompanying it within the internal auditory canal are the vestibular nerve, the facial nerve (CN VII), and the internal auditory artery. The latter provides the blood supply to the cochlea. This tiny artery originates within the intracranial vertebrobasilar arterial system and is most often a branch of the anteroinferior cerebellar artery (AICA). The cochlear nerve enters the posterior fossa at the porus acusticus and travels to the brainstem to end in the cochlear nucleus.

A lesion in one cochlear nerve will cause a unilateral hearing loss on that side. Common causes include tumors such as vestibular schwannoma (also called acoustic neuroma) and trauma with basilar skull fracture. The cochlear nerves also can be damaged by purulent meningitis, spinocerebellar degenerations, and presbycusis. CN VIII hearing loss is often accompanied by tinnitus and is typically worse at high frequencies. Sometimes the audiogram shows a tonal dip, a marked loss of sensitivity over a limited portion of the auditory spectrum.

Hearing is often lost during acoustic neuroma resection, even if the cochlear nerve remains in anatomical continuity. Sometimes this results from compromise of the internal auditory artery, but another mechanism has been elucidated by intraoperative monitoring of brainstem auditory evoked potentials, also known as auditory brainstem responses (ABRs), which can show the precise time that the hearing loss occurred. Correlation of intraoperative BAEP recordings with video recordings taken through the operating microscope showed that when hearing was lost as residual tumor was being scraped off the cochlear nerve, the instrument was usually moving from the ear toward the brainstem rather than from the brainstem toward the ear. This most likely is a consequence of the anatomy of the cochlear nerve. At the cochlear end, the fibers separate as they approach the spiral ganglion. They are weak and may be avulsed when traction is applied on the nerve toward the brainstem. The brainstem end of the nerve is mechanically much stronger. Thus, when surgical maneuvers that apply traction to the cochlear nerve, such as scraping tumor off it, are necessary, the traction should be applied toward the ear if possible.

Cochlear Nucleus

The cochlear nucleus is located on the lateral surface of the inferior cerebellar peduncle, at the rostrocaudal level of the cerebellopontine angle. The nucleus is divided into dorsal and ventral cochlear nuclei, and separate anteroventral and posteroventral subdivisions of the latter are recognized. In the cat, the cochlear nucleus can be further divided into many separate regions with distinct cellular morphologies. Cochlear nerve fibers typically bifurcate after they enter the cochlear nucleus, projecting to both dorsal and ventral cochlear nuclei. Projections to the cochlear nucleus are tonotopic (organized in an orderly fashion according to frequency, or equivalently according to the part of the cochlea where the signal originated). Multiple tonotopic maps in different cochlear nucleus regions have been identified in the cat.

Many synapses within the anteroventral cochlear nucleus contain calyxes of Held, which are very large presynaptic endings that envelop much of the postsynaptic cell. Such synapses produce a high degree of synaptic security with little jitter, which is important for accurate sound localization and also serves to maintain an afferent volley that is sufficiently coherent to produce BAEPs. Other portions of the ascending auditory pathways do not contribute to the BAEPs due to temporal dispersion of the neural activity within them. Studies of patients with neurological disease support the existence of an anatomically distinguishable brainstem neuronal subsystem for sound localization that also generates the BAEPs. BAEP abnormalities in these individuals are highly correlated with difficulty in correctly lateralizing click pairs with interaural time delays, whereas pure-tone audiometry and speech discrimination may be normal in patients with grossly abnormal BAEPs.

Lesions of one cochlear nucleus will cause a unilateral hearing loss on that side. However, strokes and tumors with effects that are largely limited to the cochlear nucleus are rare. Thus most cases of unilateral deafness in patients who are otherwise neurologically intact are due to conductive hearing loss or to disease of the inner ear or CN VIII.

Brainstem Auditory Nuclei and Tracts

The brainstem auditory pathways rostral to the cochlear nuclei include several nuclei and the interconnecting white matter pathways (**Fig. 27–1**). A prominent band

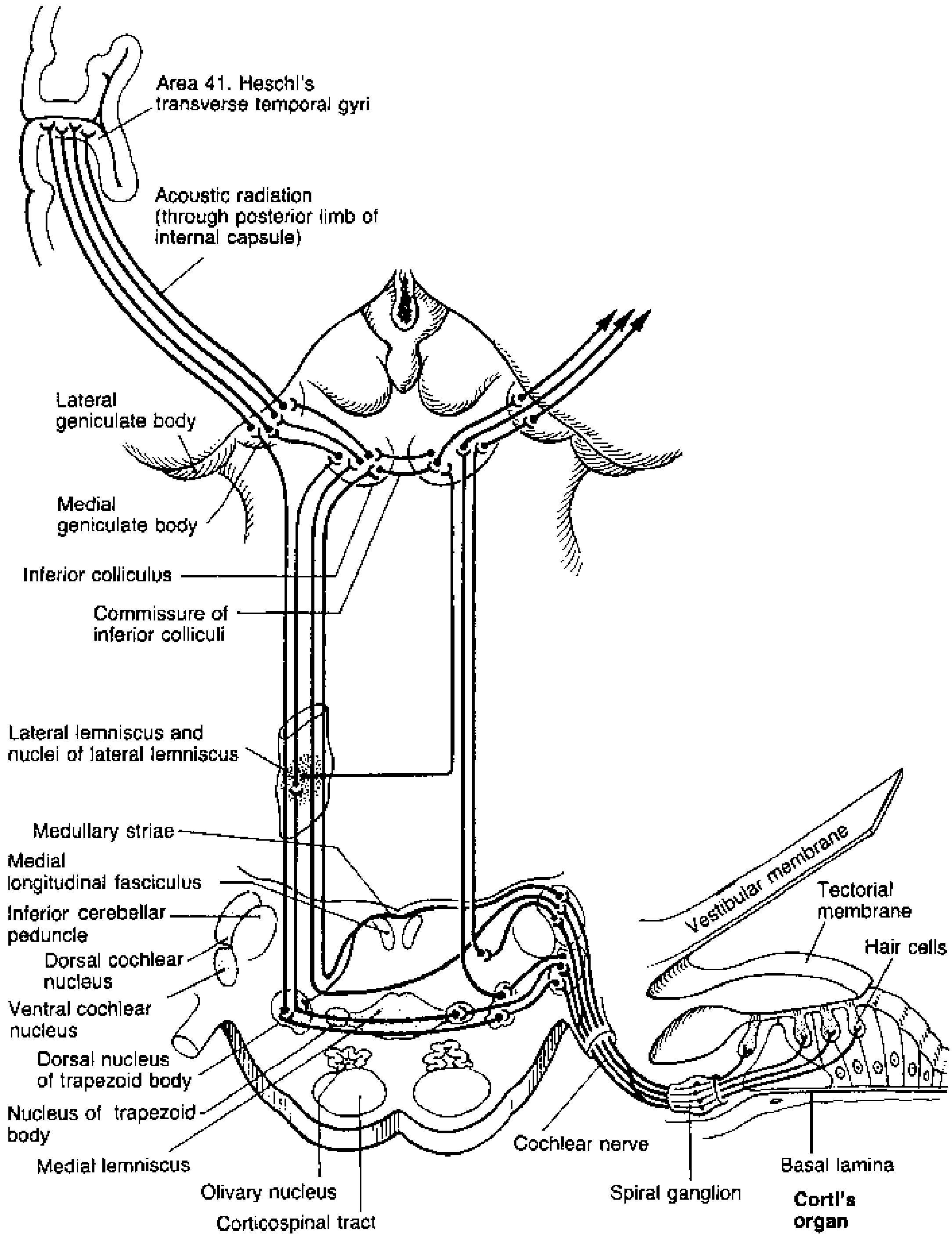


Figure 27—1 *Diagram of the anatomy of the auditory pathways, showing ascending projections from one cochlea. Projections to structures outside the classic ascending auditory pathway, such as the superior colliculus, the brainstem reticular formation, and the cerebellum are omitted. The superior olivary complex is labeled "dorsal nucleus of the trapezoid body" in this figure. (Adapted from Duus P. Topical Diagnosis in Neurology. 3rd ed. Stuttgart: Thieme Medical Publishers; 1998. Reprinted with permission.)*

of decussating auditory fibers in the lower pontine tegmentum is called the trapezoid body. Auditory nuclei within the lower pons are the medial nucleus of the trapezoid body and the superior olivary complex. The latter includes the lateral superior olive, the medial superior olive, the medial accessory olivary nucleus, and the internal and external preolivary nuclei. Despite the similarity of names, the inferior olivary nucleus of the medulla and its accompanying medial and dorsal accessory preolivary nuclei are not involved in audition. Rather, the inferior olive forms a prominent part of the circuitry of the cerebellum.

The major auditory relay nucleus at the level of the mesencephalon is the inferior colliculus, located dorsally within the quadrigeminal plate. A small band of decussating auditory fibers at this level is called the commissure of the inferior colliculus. The prominent ascending auditory white matter tract from the superior olivary complex to the inferior colliculus, coursing for the most part within the lateral brainstem tegmentum, is called the lateral lemniscus. Cellular aggregates within the lateral lemniscus comprise the nucleus of the lateral lemniscus. This is more clearly identifiable as a nucleus in cats than in primates, where it is indistinct. The next major auditory structure rostral to the inferior colliculus is the medial geniculate nucleus of the thalamus; the fiber bundle connecting these two structures is called the brachium of the inferior colliculus.

Table 27–1 Afferent brainstem auditory projections

Source	Terminations
Anteroventral cochlear nucleus (ventral acoustic stria)	Ipsilateral lateral superior olivary nucleus Bilateral medial superior olive Bilateral preolivary nuclei Contralateral medial nucleus of the trapezoid body Bilateral nucleus of the lateral lemniscus Bilateral (mostly contralateral) inferior colliculus Contralateral medial geniculate nucleus (primates)
Posteroventral cochlear nucleus (intermediate acoustic stria)	Bilateral lateral superior olivary nucleus Bilateral preolivary nuclei Contralateral medial nucleus of the trapezoid body
Dorsal cochlear nucleus (dorsal acoustic stria)	Contralateral nucleus of the lateral lemniscus Bilateral (mostly contralateral) inferior colliculus Contralateral medial geniculate nucleus (primates)

Ascending Brainstem Auditory Projections

Each cochlear nucleus sends ascending projections to both sides of the brainstem, and the afferent auditory pathways are thereafter bilateral. Thus a unilateral hearing loss implies a lesion at or distal to the cochlear nucleus. The outflow from the cochlear nucleus is organized into three fiber bundles called the acoustic striae. In order of size (number of fibers) from largest to smallest they are the ventral acoustic stria, the dorsal acoustic stria, and the intermediate acoustic stria.

The ventral acoustic stria carries the outflow from the anteroventral cochlear nucleus. Its projections are listed in **Table 27–1**. The decussating fibers of the ventral acoustic stria comprise the trapezoid body. Because each superior olivary complex receives input from both ears, a unilateral superior olivary complex lesion will not cause deafness.

The superior olivary complex is the first point in the afferent auditory system at which the same neuron receives input from both ears, permitting comparison of the binaural inputs and thus sound localization. The anatomy and physiology of the pontine auditory pathways facilitate interaural comparisons: the neurons of the medial superior olive have two tufts of dendrites at opposite sides of the cell body. Each ventral acoustic stria projects to the tuft on the same side as the cochlear nucleus from which it came. Thus each medial superior olive neuron receives geographically segregated inputs from both ears. The lateral superior olive receives a direct projection from the ipsilateral anteroventral cochlear nucleus, but the projection from the contralateral side is mediated via an interneuron originating in the medial nucleus of the trapezoid body. The synapses within the medial nucleus of the trapezoid body also utilize calyxes of Held. This is most likely because long or variable synaptic delays would compromise the accuracy of sound localization.

Although many fibers originating in the anteroventral cochlear nucleus synapse in the lower pons, others continue without synapse to the rostral pons or mesencephalon (inferior colliculus). In rhesus monkeys a few secondary afferent auditory fibers go higher, to the medial geniculate nucleus; even more do so in chimpanzees. If the phylogenetic trend continues, this may be an even more significant pathway in humans. As the neocortex assumes an increasingly prominent role in the control of behavior, a rapid pathway for transmission of auditory sensory information to primary auditory cortex is established, just as the oligosynaptic dorsal column/medial lemniscus pathways bring information to primary somatosensory cortex far more rapidly than the phylogenetically older, polysynaptic spinothalamic pathways. These "rapid transmission pathways" consist of just three neurons for both the auditory and the somatosensory systems: the primary afferent neuron, a neuron in a sensory relay nucleus (the cochlear nucleus or a dorsal column nucleus) whose axon ascends in a brainstem lemniscal tract, and a neuron in the thalamus that projects to primary sensory cortex.

The intermediate acoustic stria carries the outflow from the posterior ventral cochlear nucleus, and the dorsal acoustic stria contains the outflow from the dorsal cochlear nucleus. The projections of these striae are also listed in **Table 27–1**. Fibers from both of these striae contribute to the lateral lemnisci. Some of the axons in the dorsal acoustic striae may reach the medial geniculate nucleus without synapse.

In addition to the fibers originating in the cochlear nuclei, each lateral lemniscus contains ascending fibers originating in the superior olivary complex, the medial nucleus of the trapezoid body, and the nucleus of the lateral lemniscus. These projections predominantly remain on the same side of the brainstem as their nucleus of origin, but they contain information from both ears due to the decussations that occurred caudal to the superior olivary complex. Some afferent auditory fibers also cross to the other side of the brainstem at the level of the commissure of the inferior colliculus. Histologically, more fibers of the lateral lemniscus derive from the contralateral cochlear nucleus than the ipsilateral one. However, studies of BAEP in patients with unilateral brainstem lesions have shown that when the BAEPs are abnormal unilaterally, the abnormal BAEP is most often (though not always) obtained by stimulation of the ear ipsilateral to the lesion. Thus, whereas most brainstem ascending auditory projections are crossed, the projections in the subset of the auditory system that gives rise to the BAEPs are predominantly uncrossed.

The predominant output of the inferior colliculus is via the brachium of the inferior colliculus, which ends in the homolateral medial geniculate nucleus. The medial geniculate nucleus projects to primary auditory cortex via the auditory radiations, which course through the cerebral white matter.

In addition to the afferent auditory pathways already described, there are auditory projections to the superior colliculus, the brainstem reticular formation, and the cerebellum. These collateral pathways may be involved in reflexive movements to loud noises, which turn the head and eyes toward the sounds.

Cortical Auditory Areas

Primary auditory cortex is located in Heschl's gyri, which are on the superior surface of the temporal lobe. These gyri are hidden within the sylvian fissure and are not visible on the convexity of the brain. In the cytoarchitectural map of the human brain, primary auditory cortex corresponds to Brodmann's areas 41 and 42.

Each temporal lobe receives auditory input from both ears. Unilateral destruction of primary auditory cortex does not cause hearing loss or impaired tonal discrimination because the pathways are bilateral. Sound localization of sounds that are contralateral to the damaged hemisphere may be impaired. Bilateral auditory cortex lesions cause severe hearing loss; the patient can only hear very loud sounds and cannot discriminate tones and speech. Reported cases of this are mostly due to strokes.

For language processing, auditory information is transmitted from primary auditory cortex to other auditory cortical areas, most notably Wernicke's area. The precise anatomical location of Wernicke's area varies among individuals, but it is most commonly found in the posterior portion of the superior temporal gyrus of the language-dominant cerebral hemisphere. The left hemisphere is language-dominant in almost all right-handed individuals. The majority of left-handed individuals are also left-brain dominant for speech, though the prevalence of right-brain or mixed/bilateral language dominance in them is higher than in right-handed individuals. Damage limited to Wernicke's area typically causes a fluent aphasia, with impaired comprehension and repetition. Speech output has normal prosidy (fluency and rhythm of speech) but is marked by paraphasic errors, sometimes to the point of incomprehensibility. This is an aphasia, a deficit in the manipulation of language symbol, rather than a hearing loss per se. Studies of patients with amusia following cerebral lesions have suggested that cerebral processing of music occurs in both hemispheres, with different aspects of the musical perception processed in different cortical areas.

EFFERENT AUDITORY PATHWAYS

The ascending auditory pathways are accompanied by a complex network of descending auditory pathways that originate at almost all levels of the auditory system.

Primary auditory cortex receives its input from the medial geniculate nucleus and then sends reciprocal descending projections to that structure, as well as projections to the inferior colliculus. The various brainstem auditory nuclei are interconnected in a complex web of efferent as well as afferent fibers. For example, the inferior colliculus, the nucleus of the lateral lemniscus, and the superior olivary complex all send descending projections to the cochlear nucleus.

The cochlea itself is innervated by a major efferent projection system, the olivocochlear bundle. Each cochlea receives projections from both sides of the brainstem, with the majority originating contralaterally. Both crossed and uncrossed olivocochlear bundle fibers arise from the lower pons dorsal to the superior olivary complex, from the medial and lateral periolivary nuclei, respectively. The fibers leave the brainstem in the vestibular division of CN VIII, but cross over to the cochlear division at the commissure of Oort, deep within the internal auditory canal. They then form rich cholinergic synaptic connections on the hair cells within the cochlea. Although both excitatory and inhibitory synapses have been described, the predominant effect of the olivocochlear bundle is inhibitory; electrical stimulation of the olivocochlear bundle in experimental animals reduces the auditory nerve response to acoustic stimuli.

Efferent Auditory Pathways Mediated by Muscles

In addition to the purely neural descending systems, there are auditory feedback mechanisms mediated by contraction of muscles within the ear. The stapedius muscle is innervated by CN VII (the facial nerve). In the presence of high-intensity sound, the stapedius reflexively contracts and limits ossicular vibration, reducing the amplitude of the sounds reaching the inner ear. Contraction of the stapedius muscle modifies the mechanical compliance of the ossicular chain and thus the acoustic impedance of the tympanic membrane. The latter can be measured, and changes in acoustic impedance in response to sounds at varying intensities form the basis for acoustic impedance audiometry, which is useful for assessing hearing in young children and other uncooperative patients. Because the stapedius reflex is mediated by the facial nerve, patients with a facial palsy often complain of hyperacusis; they cannot “turn down the volume” to loud external sounds. The tensor tympani muscle, which is innervated by CN V (the trigeminal nerve), has a much smaller effect on auditory transmission; patients with CN V lesions typically do not report hyperacusis. The pathways for these myogenic efferent systems include projections from each superior olivary complex to the motor nuclei of the facial and trigeminal nerves on both sides.

AFFERENT VESTIBULAR PATHWAYS

Vestibular Receptors

The afferent vestibular system carries input from the saccule and utricle, which serve to measure head position and linear acceleration, and the three semicircular canals, which measure rotational acceleration. In each saccule and utricle, the hair cells are located in a sensory area called the macula, and their stereocilia and kinocilia are inserted into an otolithic membrane. Embedded in this membrane are otoliths or otoconia, small crystals of calcite that act as weights, shifting the membrane and thus bending the hair cell processes as the head position changes.

The three semicircular canals are arranged in an approximately mutually perpendicular fashion, so as to sense rotational acceleration along all three axes. Each semicircular canal contains a dilatation called the ampulla. The sensory hair cells are located in a region called the crista in the wall of each ampulla. Their stereocilia and kinocilia are embedded in a gelatinous plate called the cupula, which protrudes into the ampulla and is thus deflected in one direction or another depending on the direction of motion of the fluid within the semicircular canal. Fluid motion in one direction depolarizes the hair cells; fluid motion in the opposite direction hyperpolarizes them. Rotation of the head around the vertical axis produces fluid movements in the two horizontal semicircular canals that are opposite in direction with respect to the ampulla. Rotation of the face toward the left causes increased firing in the vestibular nerve fibers from the left horizontal semicircular canal and decreased firing in the vestibular nerve fibers from the right horizontal semicircular canal (**Fig. 27–2**).

Vestibular Nerve and Its Projections

Like the neurons constituting the cochlear nerve, the neurons constituting the vestibular division of CN VIII are true bipolar cells with short distal processes that receive synaptic input from the sensory hair cells and long proximal processes that travel to the brainstem within CN VIII. The cell bodies lie in the vestibular ganglion, or Scarpa’s ganglion, located within the internal auditory canal. In another similarity with the primary auditory afferent fibers, the primary vestibular fibers

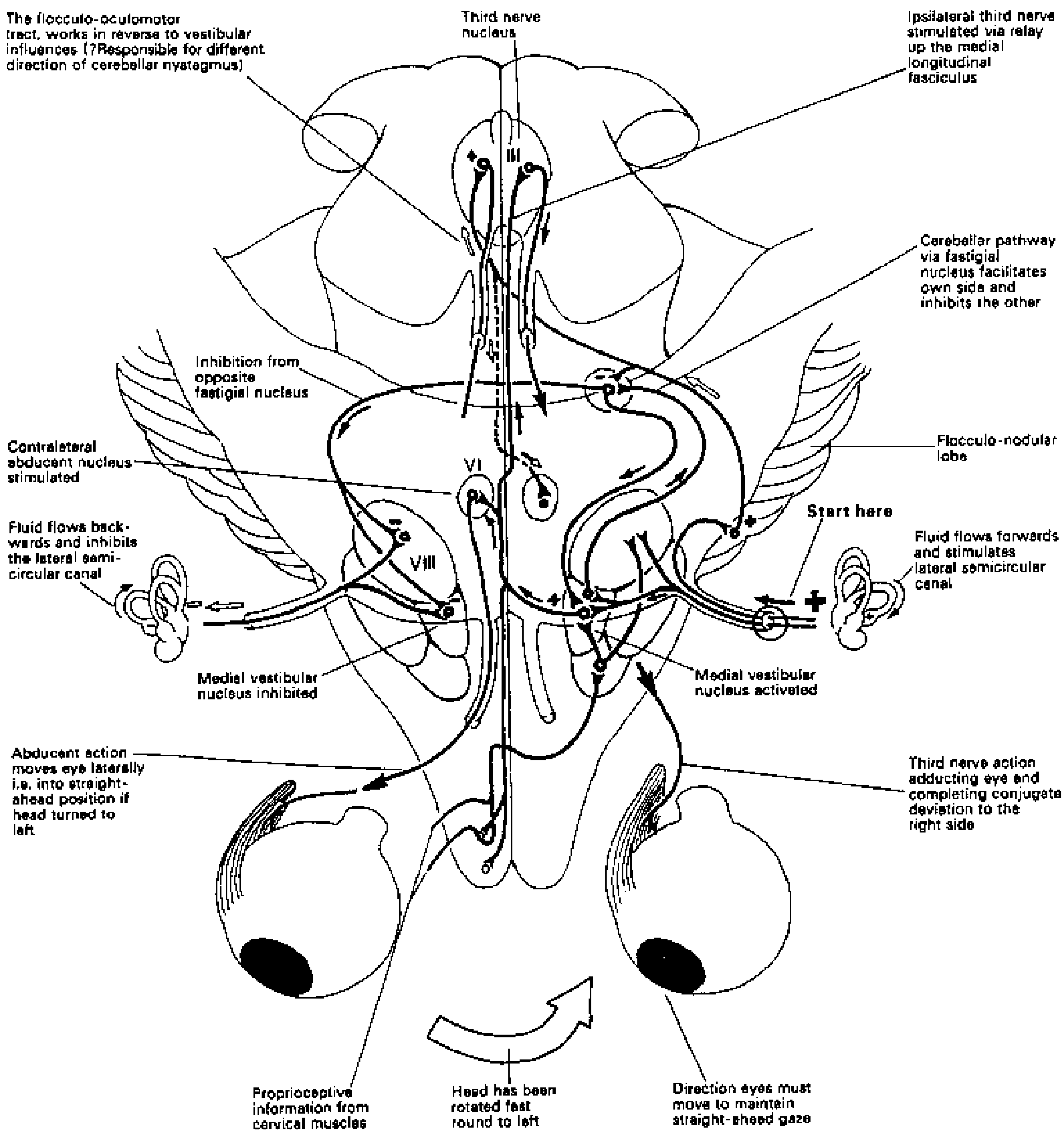


Figure 27–2 *Diagram of the vestibular pathways and the neuronal substrate for the vestibulo-ocular reflex. The changes in neuronal activity resulting from a rapid head turn to the left are indicated by solid arrows and plus signs (increased activity) or open arrows and minus signs (decreased activity). The reflexive eye movement is to the right, tending to keep the eyes fixed in space. Warm water in the left ear also causes increased activity in the left vestibular nerve, producing a slow eye deviation toward the right. If the patient is awake, there is a compensatory saccade to the left. Because nystagmus is labeled according to the direction of the fast phase, this would be nystagmus toward the same ear that was irrigated ("warm-same"). Cold water decreases vestibular nerve activity, producing changes in neuronal activity and the eye movement directions that are the reverse of those caused by warm water irrigation. The fast phase of the nystagmus will be toward the ear opposite to that which was irrigated ("cold-opposite"). (From Patten J. Neurological Differential Diagnosis. London: Harold Starke; 1977, Fig. 7.3, p. 60. Reprinted with permission.)*

bifurcate when they enter the brainstem, dividing into ascending and descending branches.

The vestibular nucleus in the dorsal pons is divided into four subdivisions: superior, inferior, lateral, and medial. Primary vestibular afferent fibers project to all four of these subdivisions. They also synapse on strands of cells located within the root fibers of the vestibular nerve; the latter have been called the interstitial nucleus of the vestibular nerve. Some vestibular nerve fibers also pass through the brainstem without synapse and enter the ipsilateral cerebellum, ending as mossy fibers within the cerebellar cortex of the flocculus, nodulus, and uvula. These latter areas have collectively been labeled the vestibulocerebellum because of their close relationships with the vestibular system.

Vestibular Nucleus and Cerebellar Projections

The secondary vestibular fibers, with cell bodies in the vestibular nuclei, project to a wide variety of CNS structures (see **Table 27–2**).

There is a large bilateral projection to the vestibulocerebellum, with reciprocal projections from that portion of the cerebellum back onto the vestibular nuclei. There are also abundant projections to the vestibular nuclei from the fastigial nuclei of the cerebellum and from the cortex of the anterior portion of the cerebellar vermis. The cerebellum projects to all divisions of the vestibular nucleus. The output of cerebellar cortex is carried entirely by Purkinje cell axons, and most of these terminate in the deep cerebellar nuclei; the direct projection from cerebellar cortex to the vestibular nuclei is the only instance in which Purkinje cell axons leave the cerebellum.

There are prominent commissural connections between the vestibular nuclei on opposite sides of the brainstem. The majority of these originate in the superior and medial vestibular nuclei (which receive input from the semicircular canals) and inhibit contralateral vestibular nucleus neurons. In contrast, the commissural projections from the areas of the vestibular nuclei that receive their inputs from the saccule and utricle have an excitatory effect on contralateral vestibular nucleus neurons.

Vestibulospinal Tract and Posturing

A group of fibers from the vestibular nuclei form the descending uncrossed vestibulospinal tract, which travels the length of the spinal cord and is strongly involved in the maintenance of posture and in vestibular-mediated postural reflexes. The vestibulospinal tract produces a strong facilitatory influence on both forelimb and hind limb extensors. This is usually modulated by the influence of descending pathways that originate more rostrally within the neuraxis. If the higher centers are destroyed, the vestibulospinal influences are unmasked, producing "decerebrate rigidity." In an animal, such as a cat, exhibiting this

Table 27–2 Projections to and from the vestibular nuclei

Nucleus	Receives projections from	Sends projections to
Superior vestibular nucleus	Cristae of the semicircular canals Cerebellum (vestibulocerebellum) Contralateral vestibular nuclei	Bilateral oculomotor nucleus (via the MLF) Ipsilateral trochlear nucleus (via the MLF) Ipsilateral interstitial nucleus of Cajal Contralateral vestibular nuclei
Lateral vestibular nucleus	Macula of the utricle Cerebellum (anterior vermis) Cerebellum (fastigial nuclei) Contralateral vestibular nuclei	Ipsilateral oculomotor nucleus (via the MLF) Ipsilateral trochlear nucleus (via the MLF) Spinal cord (via the vestibulospinal tract) Contralateral vestibular nuclei
Medial vestibular nucleus	Cristae of the semicircular canals Cerebellum (vestibulocerebellum) Contralateral vestibular nuclei Interstitial nucleus of Cajal	Bilateral abducens nucleus Contralateral oculomotor nucleus (via the MLF) Contralateral trochlear nucleus (via the MLF) Contralateral interstitial nucleus of Cajal Cervical spinal cord (via the descending MLF) Ipsilateral cerebellum (vestibulocerebellum) Contralateral vestibular nuclei
Inferior vestibular nucleus	Macula of the saccule Cerebellum (anterior vermis) Cerebellum (fastigial nuclei) Contralateral vestibular nuclei	Ipsilateral cerebellum (vestibulocerebellum) Contralateral vestibular nuclei

MLF, medial longitudinal fasciculus
Vestibulocerebellum consists of the flocculus, nodulus, and uvula.

phenomenon, destruction of the vestibular nuclei abolishes this phenomenon. Thus the presence of decerebrate rigidity implies both damage to more rostral centers (or the pathways descending from them) and at least partial survival of the vestibular nuclei.

The most rostrally originating of the descending pathways that modulate vestibulospinal activity arise from cerebral cortex. There are also descending rubrospinal tracts from the red nuclei in the mesencephalon and reticulospinal tracts from the pontine and mesencephalic reticular formation. If these brainstem areas are spared but the cortical output is interrupted, the influence of the rubrospinal and reticulospinal tracts is predominant over that of the vestibulospinal tract, producing "decorticate rigidity." This is somewhat less intense than decerebrate rigidity and also differs from it in that the arms are flexed (the legs are still extended). In general, decorticate rigidity connotes a somewhat less severe degree of brainstem damage than decerebrate rigidity. Patients who are herniating from intracranial mass lesions may progress from decorticate to decerebrate rigidity, and finally to flaccidity as the brainstem is totally destroyed.

Vestibular Control of Eye Movements

Other projections from the vestibular nuclei to the oculomotor system (**Fig. 27–2**) mediate the oculocephalic reflexes, which serve to keep the orientation of the eyes with respect to the outside world constant as the head moves, thereby stabilizing the image on the retina. The pathway for reflexive horizontal eye movements during head rotation involves projections from the vestibular nuclei to the nuclei of the third (oculomotor), fourth (trochlear), and sixth (abducens) cranial nerves. The abducens nuclei are located in the pons, close to the vestibular nuclei. The oculomotor and trochlear nuclei are located in the mesencephalon, and fibers from the vestibular nuclei reach them via the medial longitudinal fasciculus (MLF), a longitudinally oriented fiber bundle located dorsally within the brainstem tegmentum close to the midline. The MLF also extends into the cervical spinal cord, and descending MLF projections originating in the medial vestibular nucleus serve to coordinate head movements with eye movements.

Increased afferent activity from one horizontal semicircular canal causes excitation of the contralateral abducens nucleus and inhibition of the ipsilateral oculomotor nucleus, producing conjugate eye movement to the opposite side. For example, face rotation toward the left increases the activity in the left vestibular nerve, causing the eyes to rotate toward the right and thus tending to keep their orientation in space constant (see **Fig. 27–2**). Loss of vestibular input on one side, due to a lesion or dysfunction of the inner ear, vestibular nerve, or vestibular nucleus, causes conjugate eye deviation toward the side of the lesion, via a low-velocity (slow pursuit-like) eye movement. This may be followed by a corrective saccade to the opposite side; that is, away from the lesion. The sequence then repeats, producing nystagmus with the fast phase (the saccade) directed away from the vestibular lesion.

During caloric nystagmus testing, irrigation of the external ear canal with ice water produces convection currents in the horizontal semicircular canal in a direction that decreases afferent activity from that canal, mimicking a lesion on that side and producing nystagmus with the fast phase to the opposite side. Warm water produces the opposite effects; hence the mnemonic COWS (*c*old *o*pposite, *w*arm *s*ame). It is important to realize, however, that this mnemonic refers to the direction of the fast phase of the nystagmus; that is, to the corrective saccade that occurs in response to the vestibular-mediated slow eye deviation. In a comatose patient, this saccade may not occur, and the eyes will be tonically deviated toward the lesion or toward the ear that has been irrigated with ice water.

Patients with cerebral hemispheric strokes also can have tonic eye deviation, which will be toward the stroke with paresis of conjugate gaze toward the opposite side. However, the vestibular influence on the eye movement control system is powerful, and ice water calorics can cause the eyes to cross the midline in such patients, demonstrating that the lesion is above the pons.

The brainstem and cerebellar circuitry combines vestibular input from the two ears, visual input from the eyes, and somatosensory inputs from a variety of receptors (muscle spindle afferents, Golgi tendon organs, etc.) into a representation of head and body position and movement. Mismatched or conflicting information from these various inputs due to vestibular system disease can produce vertigo, nystagmus, postural unsteadiness, nausea, and vomiting. Acute labyrinthine disease may have a transient irritative phase, with increased activity in the affected vestibular nerve. However, most vestibular lesions cause decreased vestibular input ("paretic phase"), simulating head rotation away from the lesion. Patients therefore experience vertigo with a sense of rotation away from the lesion, and have horizontal nystagmus with the fast phase directed away from the lesion. In attempting to compensate for the illusory rotation, the patients tend to past-point, stagger, or fall toward the side of the lesion.

When the lesion is confined to the inner ear or CN VIII, the plasticity of the brainstem and cerebellar circuitry allows the vestibular system to adapt to the change in vestibular input, and the vertigo and nystagmus improve with time. If the loss of vestibular function is sufficiently slow, this adaptation may keep pace with it. Thus testing may reveal absent ipsilateral vestibular function in a patient with a large acoustic neuroma who had not experienced severe vestibular symptoms as the tumor slowly grew. Sacrifice of CN VIII during surgery causes no vestibular symptoms in such a patient. In contrast, patients with small acoustic neuromas may experience severe vertigo, nausea, and vomiting after surgery due to the abrupt loss of their previously relatively unimpaired vestibular input on the side of the tumor.

Additional Vestibular and Eye Movement–Related Projections

The interstitial nucleus of Cajal, a group of cells located within the mesencephalic MLF, is also involved in vestibular processing. It receives input from the ipsilateral superior vestibular nucleus and the contralateral medial vestibular nucleus and sends projections to the cervical spinal cord; these projections travel through the descending MLF along with the fibers originating in the vestibular nuclei.

Excitatory ascending projections from the abducens nucleus, which decussate and then travel through the brainstem MLF to the contralateral oculomotor nucleus, serve to coordinate conjugate horizontal eye movements by causing simultaneous contraction of the lateral rectus (innervated by the abducens nerve) on one side and the medial rectus (innervated by the oculomotor nerve) on the opposite side. An MLF lesion blocks the projection to the oculomotor nucleus on that side, causing an "intranuclear ophthalmoplegia" with failure of adduction of the ipsilateral eye on attempted conjugate gaze to the opposite side. This is accompanied by nystagmus of the abducting eye; the pathophysiology of the nystagmus is unclear. Attempted ocular convergence can still cause the eyes to adduct, demonstrating that the oculomotor nuclei themselves and their connections to the medial rectus muscles are intact and that convergence is mediated by brainstem pathways distinct from the MLF.

Ascending projections from the vestibular nuclei to the thalamus terminate predominantly in the nucleus ventralis posterolateralis (VPL). In contrast to the auditory and visual systems, which have separate thalamic relay nuclei (the medial and lateral geniculate nuclei, respectively), there does not appear to be a separate vestibular thalamic nucleus; thalamic areas that receive vestibular input also respond to somatosensory input. Similarly, some areas of cerebral cortex respond to both somatosensory and vestibular input. However, several lines of evidence (clinical-pathological correlations in patients with strokes, vestibular stimulation studies in human subjects, and homologies with the vestibular cortex of subhuman primates) identify the posterior insula as the predominant vestibular representation within human cerebral cortex.

EFFERENT VESTIBULAR PATHWAYS

The major projections to the vestibular nuclei from other brain regions arise from the cerebellum and from the vestibular nuclei on the opposite side, as already described. There are also descending pathways from vestibular cortex to the vestibular nuclei.

Like the auditory system, the vestibular system also contains an efferent pathway that synapses on the sensory hair cells within the inner ear. It is thought to modulate the dynamic range of the vestibular afferents to match expected acceleration. The efferent fibers arise from cell bodies located lateral to the abducens nuclei, travel within the vestibular nerves, and make excitatory synapses on the hair cells of the cristae and maculae. Like the olivocochlear bundle, the efferent vestibular projections include both crossed and uncrossed pathways. Also, the synapses on the hair cells are cholinergic for both the efferent auditory projections and the efferent vestibular projections to the inner ear.

Suggested Readings

Parent A. Carpenter's Human Neuroanatomy. 9th ed. Baltimore: Williams & Wilkins; 1996

Patten J. Neurological Differential Diagnosis. 2nd ed. London: Springer; 1996

Strominger NL. The origins, course and distribution of the dorsal and intermediate acoustic striae in the rhesus monkey. J Comp Neurol 1973;147:209–234

Strominger NL, Nelson LR, Dougherty WJ. Second-order auditory pathways in the chimpanzee. J Comp Neurol 1977;172: 349–366

Strominger NL, Strominger AI. Ascending brain stem projections of the anteroventral cochlear nucleus in the rhesus monkey. J Comp Neurol 1971;143:217–242

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Unilateral cochlear stimulation will result in afferent neuronal activity within
   A. One auditory nerve, one superior olivary complex, one inferior colliculus, and one auditory cortex
   B. One auditory nerve, one superior olivary complex, one inferior colliculus, and both auditory cortices
   C. One auditory nerve, one superior olivary complex, both inferior colliculi, and both auditory cortices
   D. One auditory nerve, both superior olivary complexes, both inferior colliculi, and both auditory cortices
2. Contraction of the stapedius muscle
   A. Increases the acoustic impedance of the tympanic membrane
   B. Is mediated by the fifth cranial nerve
   C. Can result in hyperacusis
   D. Increases the transmission of sound through the ossicles
3. A patient suddenly loses function in the right labyrinth. What are the symptoms and signs?
   A. The patient will experience vertigo, with the feeling that he or she is rotating from right to left. The patient will have nystagmus, with the fast phase to the left, and will tend to fall to the right when walking.
   B. The patient will experience vertigo, with the feeling that he or she is rotating from right to left. The patient will have nystagmus, with the fast phase to the right, and will tend to fall to the right when walking.
   C. The patient will experience vertigo, with the feeling that he or she is rotating from left to right. The patient will have nystagmus, with the fast phase to the left, and will tend to fall to the left when walking.
   D. The patient will experience vertigo, with the feeling that he or she is rotating from left to right. The patient will have nystagmus, with the fast phase to the right, and will tend to fall to the left when walking.
4. An unconscious patient whose labyrinth, cranial nerves, and brainstem are intact is positioned supine, with the head elevated 30 degrees. What will happen when ice water is instilled into the left external auditory canal?
   A. Afferent activity in the left vestibular nerve will increase, and the eyes will deviate tonically to the left.
   B. Afferent activity in the left vestibular nerve will increase, and the eyes will deviate tonically to the right.
   C. Afferent activity in the left vestibular nerve will decrease, and the eyes will deviate tonically to the left.
   D. Afferent activity in the left vestibular nerve will decrease, and the eyes will deviate tonically to the right.

Chapter 28

ASSESSMENT OF CENTRAL AUDITORY FUNCTION

Philippe P. Lefebvre and Alan D. Legatt

Many practicing otolaryngologists are unfamiliar with central auditory processing and its abnormalities. The usual auditory function tests assess the peripheral organ (including both the middle ear and the inner ear) and the cochlear nerve. Some patients in whom the results of these tests, including a speech discrimination test, show normal hearing still complain of listening and understanding disabilities. These patients have conditions that have been called auditory perceptual disorders, auditory processing disorders, or, more commonly, central auditory processing disorders. Central auditory function testing is used to identify and characterize abnormalities of central auditory processing in these patients.

Central auditory function tests generally are used to characterize abnormalities of function rather than to diagnose macroscopic structural lesions affecting the central auditory pathways; lesions can be identified using sophisticated imaging techniques such as magnetic resonance imaging (MRI) and computed tomographic (CT) scans. Brainstem auditory evoked potentials (BAEPs), also called auditory brainstem responses (ABRs), can help to localize abnormalities anatomically. In children presenting with delayed or abnormal language acquisition, central auditory function testing provides information about the level of maturation of the central auditory pathways. Documentation of a central auditory processing disorder allows caregivers to choose the best and most appropriate remediation strategy for the child. This chapter provides information about central auditory function testing and its results in several pathological conditions.

DEFINITIONS

Central auditory processes are the neuronal mechanisms responsible for the following behavioral phenomena: sound localization and lateralization, auditory discrimination, auditory pattern recognition, temporal aspects of audition (including temporal resolution, temporal masking, temporal integration, and temporal ordering), auditory performance with competing acoustic signals, and auditory performance with degraded acoustic signals (American Speech Language Hearing Association, 1996). Central auditory processing disorders can be defined as abnormalities of the basic processes involved in understanding spoken language in the absence of dysfunction of the peripheral auditory system lesion (ear and cochlear nerve). They manifest as a deficit in information processing of the audible signal and/or as an impaired ability to discriminate, remember, recognize, and comprehend information presented to normal ears. The neuronal abnormalities that cause these disorders, therefore, must be localized between the cochlear nucleus and auditory areas of cerebral cortex.

NEUROPHYSIOLOGICAL BASIS OF CENTRAL AUDITORY FUNCTION

Several principles regarding the treatment of the auditory messages by the central nervous system (CNS) are summarized here.

- Channel separation: Information about an acoustic signal delivered to one ear is transferred to, and maintained in, the auditory cortex in a manner that keeps it distinct from information about a signal delivered to the contralateral ear.
- Binaural fusion: If a unique message is separated into two bands and if these two bands are delivered simultaneously to both ears, a fusion occurs at the level of the brainstem, and the subject will perceive only one message.
- Bilateral pathways: In normal subjects, information about an auditory signal delivered to one ear travels through both direct and crossed auditory pathways and reaches both temporal lobes. Ascending auditory fibers cross the midline at multiple levels within the brainstem. Furthermore, at the cerebral level, auditory information can cross from one hemisphere to the other through the corpus callosum, and perhaps through the anterior commissure as well.
- Cerebral dominance: The left cerebral hemisphere is "dominant" with respect to the perception of language in the majority of the population. However, some elements of the auditory message are processed in the "nondominant" right hemisphere.

AUDIOLOGICAL TESTS AND CENTRAL AUDITORY PROCESSING

Pure-Tone Audiometry

When testing central auditory function, pure-tone audiometry should be performed to identify any peripheral auditory dysfunction. Pure-tone audiometry usually will not detect a lesion of the central auditory pathways and will not be modified by a lesion above the inferior colliculus because intensity and frequency discrimination are performed at a level below the inferior colliculus.

Speech Audiometry

Speech audiometry assesses the ability of subjects to hear and understand the spoken word. A list of monosyllabic or disyllabic words is presented through prerecorded tapes to the subject at different intensities. The list consists of either phonetically balanced or isophonemic monosyllabic or spondaic (equally accented disyllabic) words. Phonetically balanced lists are those in which each phoneme in the list appears in proportion to its frequency of occurrence in natural language. Isophonemic lists are those in which a phoneme occurs once only in each list. The number of words in each list is usually 10. Two parameters are measured. The speech detection threshold (SDT) is the intensity at which 50% of the words are detected but not understood. The speech reception threshold (SRT) is the intensity at which 50% of the words are correctly reported.

The value of speech audiometry lies in the fact that many retrocochlear lesions have a greater effect on speech comprehension than on the pure-tone audiogram. In contrast to patients with a sensory hearing loss (in whom the predominant pathology is hair cell loss, with or without a loss of cochlear nerve fibers), the speech audiogram of a subject presenting with a retrocochlear lesion is often worse than anticipated from the mean hearing loss of the pure-tone audiogram. Speech audiometry frequently is severely abnormal in the affected ear with unilateral cochlear lesions, but results are usually not so asymmetrical in patients with central lesions.

Masked Speech Audiometry

There are many ways of performing speech audiometry to attempt to detect lesions or abnormalities of auditory processing in the brainstem or cerebral cortex, including the use of distortions of spoken words and periodic interruptions of speech. To be useful in the clinic, however, a simple method is necessary. Masked speech audiometry is of some value. Basically, the technique involves the presentation of words simultaneously with noise to one ear and comparing the results with those obtained when a list of words is presented in the absence of noise. Most patients with lesions of the central auditory pathways have a markedly reduced ability to understand the words in the presence of noise. Interestingly, the side of poorest performance is not always consistent with that of the lesion; in some patients, the abnormalities are bilateral. Furthermore, the rostrocaudal level of the lesion (e.g., pons or mesencephalon) does not correlate with the degree of abnormality of the masked speech audiogram.

Masking Level Differences

The masking level differences (MLDs) test is a valuable test for the assessment of central auditory dysfunction. In essence, the test consists of the presentation of a pulsed 500 Hz tone to both ears (binaurally), in the presence of continuous broad-band noise presented at 60 dB. The intensity of the tone, which is pulsed at a rate of 200 msec on/200 msec off, is varied until the patient indicates that the tone is perceived. Two test conditions are evaluated. In the homophasic condition, the stimulus and the noise are presented to both ears in phase, and the detection threshold of the tone is determined. In the antiphasic condition, the signal is 180° out of phase between the two ears, whereas the noise remains in phase. The threshold obtained in antiphasic condition is subtracted from the threshold obtained in the homophasic condition to determine the amount of release from masking in dB. Masking level differences of less than 6 dB are considered abnormal in adult patients. The MLD test is sensitive to lesions in the lower brainstem, but it is largely unaffected by rostral brainstem or cortical lesions. Furthermore, there is a close correspondence between MLD and BAEP test results: patients with abnormalities in waves I, II, or III of the BAEP demonstrate little or no release from masking, whereas patients with abnormalities of waves IV and V show normal MLD results.

Dichotic Speech Testing

Dichotic speech tests consist of the simultaneous presentation of different stimuli to each ear. The interpretation of the dichotic tests is based on the model developed by Kimura, which has the following premises: (1) the contralateral auditory pathways in humans are more numerous and robust than the ipsilateral pathways; (2) when monaural input is presented to the system, either pathway is capable of initiating and conducting the appropriate neural response; and (3) in dichotic situations, the weaker ipsilateral pathway is suppressed and the stronger contralateral pathways remain active. Hence, if one hemisphere is compromised, a deficit would be anticipated.

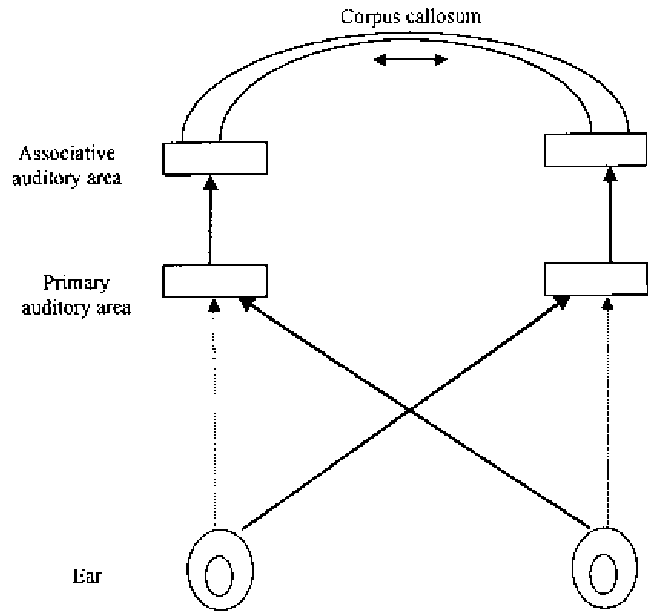


Figure 28–1 *Central auditory pathways in dichotic hearing. (From Sparks R, Goodglass H, Nickel B. Ipsilateral versus contralateral extinction in dichotic listening resulting from hemisphere lesions. Cortex 1970;6:249–260. Reprinted with permission.)*

Two tests are widely used to evaluate the patients: the dichotic digits test and the staggered spondaic test. In the dichotic digits test, digits are presented to each ear in a dichotic fashion at a comfortable intensity level (usually 50 or 60 dB), and the patients are asked to repeat all digits heard. The digit dichotic test is fairly sensitive to intracranial lesions. In the staggered spondaic words test, developed by Katz (1977), 40 pairs of spondee words are presented to the patient in an overlapping but staggered fashion, in both competing and noncompeting conditions. These tests tend to reveal the site of the lesion along the central auditory pathways, in particular temporal lobe lesions or interhemispheric lesions (**Fig. 28–1**).

ELECTROPHYSIOLOGICAL TESTS AND CENTRAL AUDITORY PROCESSING

Acoustic Reflex

The efferent auditory system provides feedback control of the volume of sounds reaching the cochlea. Loud sounds cause reflexive contraction of the stapedius muscle, which limits ossicular vibration.

Because the tympanic membrane is connected to the ossicles, stapedius contraction modifies the acoustic impedance of the tympanic membrane and thus the acoustic impedance of the middle ear. This latter can be measured ("impedance audiometry"), and changes in it in response to sounds provide an assessment of the acoustic reflex.

The acoustic reflex is not considered a strong measure of central auditory function. However, the reflex may be disturbed by a lesion situated in the lower brainstem (the pons); the anatomical pathway subserving the acoustic reflex involves projections from the ventral cochlear nuclei to the superior olivary complexes and from there to the facial motor nuclei. Because of this, the acoustic reflex can be valuable in assessing central auditory integrity in children and adults.

Evoked Otoacoustic Emissions

Otoacoustic emissions are low-level acoustic signals that are generated by the cochlea, both spontaneously and in response to auditory stimulation. The latter include (1) transiently evoked otoacoustic emissions (TEOAEs), which are produced in response to short-duration signals such as clicks and tone bursts; (2) distortion product otoacoustic emissions (DPOAEs), which are evoked when stimuli of two different frequencies are presented; and (3) stimulus-frequency otoacoustic emissions (SFOAEs), which are generated by continuous pure-tone stimuli that vary slowly in frequency. The TEOAEs and the DPOAEs are most widely used for clinical testing.

Although otoacoustic emissions should be strictly normal in central auditory processing disorders, they can be valuable in evaluating patients suspected of having such disorders to prove the normal function of the cochlea; it is known that cochlear abnormalities can be present in the face of a normal pure-tone audiogram. The measurement of otoacoustic emissions can provide an index of the integrity of the olivocochlear bundle, which originates in the region of the superior olivary complex and terminates at the base of the outer hair cells (medial system) and at the nerve fibers at the base of the inner hair cells (lateral system). Otoacoustic emissions are influenced by stimulating the olivocochlear bundle. When noise is presented to the contralateral ear, otoacoustic emissions are suppressed in amplitude by several decibels. Because the olivocochlear bundle appears to play a role in the ability to hear sounds in the presence of noise, it may be important to know whether the olivocochlear bundle is functioning normally in patients presenting with central auditory processing disorders.

Brainstem Auditory Evoked Potentials

BAEPs are the electrical signals produced by the infratentorial auditory system in response to transient auditory stimuli such as clicks or brief tone pips. Stimuli are presented monaurally, and masking noise is presented to the contralateral ear. The responses are recorded between the vertex and the ear or mastoid ipsilateral to the stimulated ear; additional recording channels are often used to clarify components or assist in their identification. The BAEP peaks, which have latencies of less than 10 msec, are typically labeled with roman numerals according to the convention of Jewett and Williston (**Fig. 28–2**); waves IV and V are often fused into a IV–V complex. Component amplitudes vary from subject to subject, but peak latencies are highly consistent across subjects. Thus the interpretation of BAEPs is predominantly based on measurements of absolute peak latencies, interpeak intervals, and the right-left differences of these measures; the IV–V: I amplitude ratio is also useful in identifying neurological abnormalities.

As the stimulus intensity is reduced, BAEP components increase in latency and decrease in amplitude, and eventually disappear. Thus BAEPs can be used to estimate hearing thresholds. Frequency-specific stimuli (e.g., tone pips) and frequency-specific masking (e.g., notched noise) can be used to measure thresholds at various frequencies and produce audiograms that are similar to behavioral audiograms obtained in the same subject. However, BAEPs are most useful for the assessment of retrocochlear abnormalities.

Wave I is generated in the distal (i.e., at the cochlear end) cranial nerve (CN) VIII, and may be preserved in lesions of the proximal CN VIII. It is identical to the electrocochleographic N1, the first peak of the CN VIII compound action potential. A delay in wave I indicates peripheral auditory dysfunction, such as a conductive or cochlear hearing loss. All of the subsequent BAEP components are the composites of contributions from multiple generators (**Fig. 28–2**). For example, wave II includes contributions both from the cochlear nucleus and from the second volley in the distal CN VIII (the electrocochleographic N2). However, in clinical-pathological correlations and localization of lesions with BAEPs, wave III predominantly reflects activity at the level of the lower pons, and wave V predominantly reflects activity at the level of the mesencephalon. Clinical

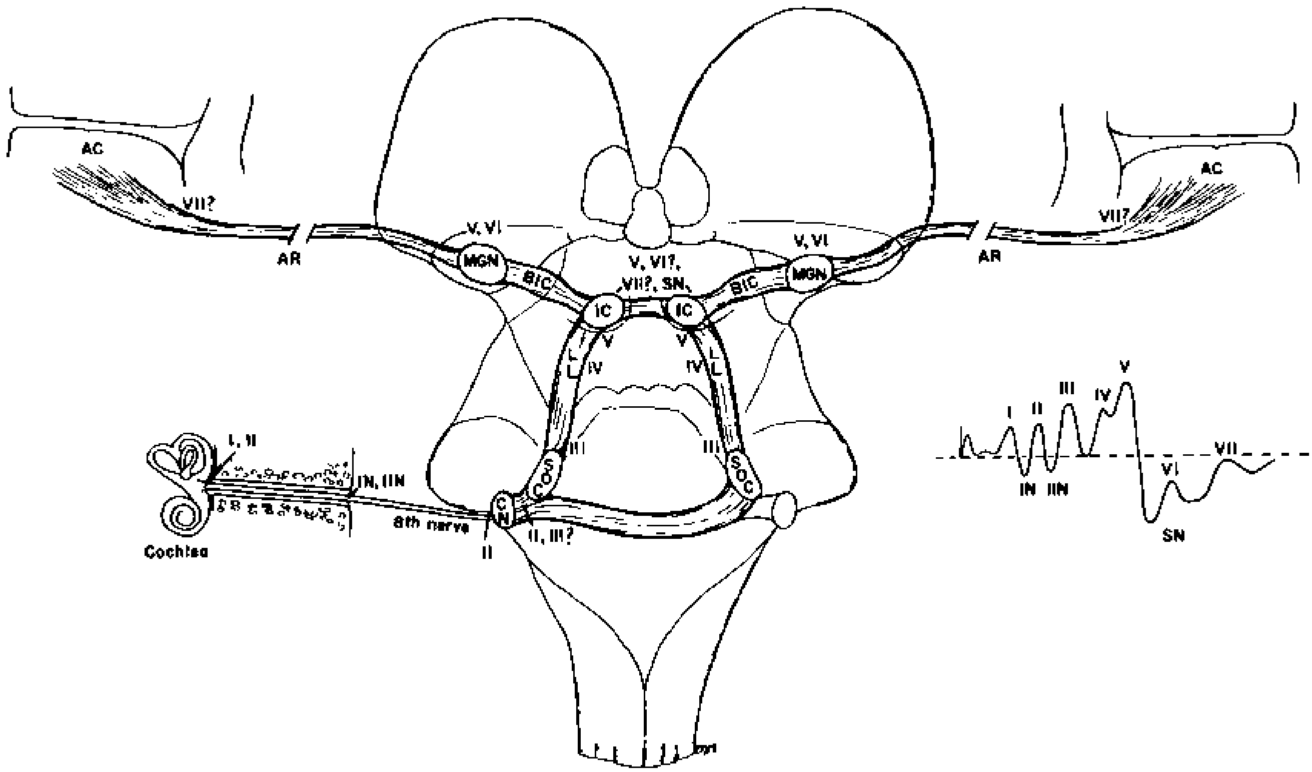


Figure 28–2 Diagram showing the probable generators of the human brainstem auditory evoked potentials (BAEPs). AC, auditory cortex; AR, auditory radiations; BIC, brachium of the inferior colliculus; CN, cochlear nucleus; IC, inferior colliculus; LL, lateral lemniscus; MGN, medial geniculate nucleus; SN, slow negativity after wave V; SOC, superior olivary complex. (From Legatt AD, Arezzo JC, Vaughan HG Jr. The anatomic and physiologic bases of brain stem auditory evoked potentials. Neurol Clin 1988;6:681–704, Fig. 12, p. 698. Reprinted with permission.)

interpretation of BAEPs is based predominantly on waves I, III, and V; the other components are variable and are occasionally not identifiable in normal subjects. BAEPs cannot be used to assess the auditory pathways rostral to the mesencephalon. For example, patients with bilateral temporal lobe infarctions involving auditory cortex may be deaf yet have completely normal BAEPs.

A unilateral cochlear, CN VIII, or cochlear nucleus lesion will cause unilateral BAEP abnormalities affecting the waveforms to stimulation of the ear ipsilateral to the lesion. The ascending auditory pathways become bilateral at the level of the superior olivary complex. Although more ascending fibers rostral to this level are crossed than are uncrossed, the subset of ascending auditory neurons generating the BAEPs are predominantly ipsilaterally driven because unilateral brainstem lesions that produce unilateral abnormalities involving either the I–III or the III–V interpeak interval usually do so upon stimulation of the ear ipsilateral to the lesion.

BAEPs are highly sensitive (>95% sensitivity) in the detection of tumors of CN VIII. Small tumors may prolong the I–III interpeak interval or cause loss of waves III and V. As the tumors grow, they can compress the internal auditory artery, which originates in the intracranial circulation (usually as a branch of the anterior inferior cerebellar artery) and courses through the internal auditory canal next to CN VIII to supply blood to the cochlea. This can prolong the latency of wave I and even cause the loss of wave I and all subsequent BAEP components, due to cochlear ischemia or infarction. With very large tumors, the III–V interpeak interval to stimulation of the contralateral ear may become prolonged, reflecting brainstem compression. BAEPs are used for intraoperative monitoring during resection of CN VIII tumors, as well as during other temporal bone or posterior fossa operations in which the ear, CN VIII, or brainstem are at risk.

BAEPs are used as a diagnostic test in patients suspected of having multiple sclerosis, though their sensitivity is less than that of visual evoked potentials. Demyelination results in a slowing of neural conductions and thus prolongation of BAEP peak latencies and interpeak intervals, which can be recognized at a point when the myelin damage is subclinical. BAEPs are also highly sensitive in the detection of intrinsic brainstem tumors, such as gliomas.

BAEPs can detect unilateral brainstem lesions affecting the auditory pathways in patients with normal audiograms; these patients have no hearing loss because the ascending auditory pathways are bilateral. BAEPs can also detect central auditory processing abnormalities in patients with bilateral brainstem disease that slows but does not interrupt neural conductions; these patients may have normal hearing because the information that the sound has occurred, though delayed, still reaches the auditory cortex.

CENTRAL AUDITORY PROCESSING DISORDERS

Central auditory processing disorders can be defined as a dysfunction of the basic processes involved in understanding the spoken language in the absence of a peripheral auditory system lesion. Many clinicians believe that children with learning disabilities are the only population in whom central auditory testing is appropriate. However, it may be of value in many other clinical situations.

Patients with degenerative neurological diseases that may affect the auditory pathways are another population that deserves central auditory evaluation. The most common of these diseases that has an auditory correlate is multiple sclerosis, in which the pathophysiology involves demyelination of axons within the central nervous system. Abundant evidence indicates that individuals with multiple sclerosis can have auditory deficits, primarily when the auditory pathways are involved. Disorders of central auditory processing, such as impaired sound localization, have been found in multiple sclerosis patients who have normal audiograms. BAEPs demonstrate the central dysfunction in these subjects, and the degree of BAEP abnormality correlates with the degree of impairment of sound localization.

The absolute latency of wave I of the BAEP is sometimes called the "peripheral transmission time," and the I–V interpeak interval is labeled the "central transmission time." The latter might seem inappropriate because wave I originates in the most distal portion of CN VIII, and conduction along the cranial nerve is thus included within the "central transmission time." However, the cochlear nerve axons are ensheathed by CNS-type myelin, produced by oligodendrogliocytes, along most of CN VIII; the transition to peripheral nervous system–type myelin, produced by Schwann cells, occurs close to the cochlea. CN VIII is therefore vulnerable to diseases that affect CNS myelin, such as multiple sclerosis.

Neurological degenerative diseases other than multiple sclerosis also have been shown to involve the auditory system; for example: Charcot-Marie-Tooth disease, Alzheimer's disease, olivopontocerebellar degeneration, Friedreich's ataxia, Parkinson's disease, and various leukodystrophies. However, these diseases are not as common as multiple sclerosis, nor have they been studied as much from an auditory perspective.

Patients with seizure disorders with epileptogenic foci at or near the auditory areas of the cerebrum also are candidates for central auditory assessment. Epilepsy or the causative lesions can cause dysfunction of the central auditory nervous system. In addition, two surgical primary treatments for intractable epilepsy, corpus callosotomy (split brain) and resection of the seizure focus, may affect central auditory processing. The split brain procedure and its effect on auditory processing have been well studied. Also well studied from an audiological point of view is the effect of anterior temporal lobectomy, which is a common surgical treatment for intractable focal epilepsy. Preservation of the temporal lobe speech centers is important during temporal lobectomy in the language dominant hemisphere.

Patients with mass lesions of the central auditory nervous system also are candidates for central auditory evaluation. By using central auditory tests, insight may be gained as to how much the auditory system is compromised and what may be the functional consequences of these lesions or their surgical removal.

Patients who have suffered head trauma often have damage to the central and/or peripheral auditory systems. Sometimes central auditory function tests can provide insight into the nature of the recovery or the lack thereof in these patients. Unfortunately, central auditory function tests are seldom employed in these patients, even though extensive neuropsychological testing is often performed in them.

Patients who wear, or are candidates for, hearing aids and who do not do well with amplification may require central auditory assessment. In many cases, hearing aids are not effective because of some compromise of central auditory processing. Often, these patients have a history of CNS pathology, which has affected their central auditory processing. Also, the great majority of hearing aids users are older adults, who are at risk for central auditory processing abnormalities. Management of patients who do poorly with amplification can be improved if a central deficit is defined. Peripheral auditory function may be symmetrical, or one ear may be better, but a central evaluation may indicate an entirely different result in regard to ear symmetry. This information may have implications for the management of the patient. Some patients suffer from a binaural interference phenomenon, in which extremely poor speech recognition on one side contaminates good performance from the other side. In this condition, binaural amplification may be worse than a single hearing aid.

Patients who complain about difficulty hearing in noisy environments and have normal audiograms suffer from obscure auditory dysfunction, and their central auditory function should be evaluated. These patients are most often adults without underlying disorders. Obscure auditory function represents a relatively large population of patients who, until recently, were dismissed by clinicians as having no hearing problems.

Suggested Readings

American Speech Language Hearing Association. Central auditory processing: current status of research and implications for clinical practice. Am J Audiol 1996;5:41–54

Katz J. The staggered spondaic word test. In: Keith R, ed. Central Auditory Dysfunction. New York: Grune & Stratton; 1977

Legatt AD. Brainstem auditory evoked potentials: methodology, interpretation, and clinical application. In: Aminoff MJ, ed. Electrodiagnosis in Clinical Neurology. 4th ed. New York: Churchill Livingstone; 1999:451–484

Probst R, Harris FP. Otoacoustic emissions. In: Alford BR, Jerger J, Jenkins HA, eds. Electrophysiologic Evaluation in Otolaryngology. Basel: Karger; 1997:182–204. Advances in Otorhinolaryngology; vol 53

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. A patient is felt to have a purely central auditory processing disorder. Which of the following findings would be inconsistent with that diagnosis?
   A. Impaired ability to lateralize sounds in space
   B. A high-frequency hearing loss
   C. A masking level difference of less than 6 dB
   D. An elevated speech reception threshold in the presence of a normal pure-tone audiogram
2. Otoacoustic emissions
   A. Can be recorded following a continuous pure-tone stimulus but not following a click stimulus
   B. Are usually abnormal in patients with central auditory processing disorders
   C. Are affected by lesions of the inferior colliculus
   D. Are affected by stimulation of the olivocochlear bundle
3. Brainstem auditory evoked potentials (BAEPs) can detect abnormalities
   A. Within cranial nerve (CN) VIII, but not within the pons, medial geniculate, or auditory cortex
   B. Within CN VIII or the pons, but not within the medial geniculate or auditory cortex
   C. Within CN VIII, the pons, or the medial geniculate, but not within the auditory cortex
   D. Within CN VIII, the pons, the medial geniculate, or the auditory cortex
4. In multiple sclerosis, demyelination
   A. Can involve CN VIII
   B. Prolongs BAEP peak latencies but not interpeak intervals
   C. Can increase conduction velocities within the central auditory pathways
   D. Affects the audiogram and sound localization abilities equally

Chapter 29
LANGUAGE AND THE PLASTIC BRAIN

ROBERT J. RUBEN

There is but one area of evolutionary specialization that has enabled the Darwinian success of humans: language. The foundation of our present evolutionary success is the ability to increase our knowledge through our use of language. All histories, archives, and knowledge bases are a result of our linguistic abilities, and there appears to be no other life form that is so well endowed with our particular linguistic traits. What are the biological characteristics that underlie this remarkable evolutionary product?

The conceptualization of language has undergone its own development. The ancients saw language as a fixed human characteristic and felt that it was an inherent function, and furthermore that there was a "primary" language. Herodotus, in the 5th century BC, referred to a language experiment of sorts conducted by Pharaoh Psammetichus, who had placed two children who were being nursed by goats in isolation for 2 years. The children's vocalizations were noted at the end of those 2 years. Their utterance was the sound "Becos," which was interpreted as being the Phrygian word for bread. Psammetichus was reportedly disappointed because he had hoped that Egyptian would have been the children's "primary" language. More likely the children were using the sounds that they had heard, which was the bleating of their "nurses," the goats. One could characterize the pharaoh's interpretation of the study as the extreme example that language was all nature—or an intrinsic human ability; for him, the bleating of the goats had no effect on the children.

The intrinsic nature of language remained the dominant idea until the beginning of the 19th century. It is surprising that awareness of the lack of language in the congenitally deaf and their occasional habilitation, beginning in the 17th century, did not result in consideration of the nurture/extrinsic contribution to the establishment of a person's language. Two case reports, that of Itard in 1801 and Wardrop in 1813, concern themselves with a few years of life and its effect on the linguistic abilities of the child. Wardrop's patient, James, illustrate the effect of deprivation of auditory linguistic input on a child who had adventitious linguistic exposure—sight, touch, and so on. Although deaf from birth (in retrospect, James was probably a child with congenital rubella), James was able to have emotion and rudimentary communication skills. Wardrop's study illustrates the tenet of an intrinsic linguistic mechanism that needs very little extrinsic input to achieve a modicum of language. Itard's patient, Victor, had normal hearing but was probably a victim of social deprivation: he was found in a forest and was assumed to have been raised by animals. After much work, the boy attained little more than rudimentary linguistic ability. Itard's study illustrates that there is a need for extrinsic linguistic input for the development of a typical communication system. These two reports were instrumental in the awareness that a person's language was dependent on both intrinsic and extrinsic factors and the beginning of the concept of a critical/sensitive period in language development.

CRITICAL PERIODS

The need for intrinsic brain mechanisms and external linguistic input and the time frame of critical periods have been further quantified by several controlled and/or prospective human studies performed over the past 3 decades. The acquisition of language appears to be a progression that first involves phonology (the sound of words), followed by semantics (the meaning of words) and syntax (the rules of grammar).

The ability of the person to select from the cacophony of environmental sounds that are salient for language is a developmental process that begins at least by the sixth month of fetal life and approaches maturity by the end of the first postpartum year of life. The fetus develops in a sound-attenuated environment—the womb—with audible frequencies between 30 and 40 dB. The ability of the fetus to react to sound has been demonstrated by observing through ultrasound eye blinks to pulsed tones at 28 weeks gestation, as well as by noting electrocardiogram changes with sound stimuli. This gives rise to the question, Do such physiological responses in any way reflect, or themselves create, a shaping of the fetal central nervous system (CNS) with regard to preferential sound perception for the language to which they are exposed in utero? Simply stated, Do the mother's voice and native language sounds have any advantage over all other sounds at birth? The answer is yes. The newborn prefers the mother's voice to other female voices and the mother's voice to a male voice. It has further been shown that the newborn will recognize sounds and patterns heard in utero. Other studies (Jusczyk et al. [1988] and Mehler et al. [1988]) have demonstrated that infants prefer the unique sounds of their mother's language to foreign languages. One study was performed on 2-day-old infants and the other at 2 months of age. Collectively, these data indicate that there is intrauterine phonemic learning.

During the first year of life, the infant evolves from a relatively poorly differentiating language receptor to one exquisitely tuned to the unique sounds—phonemes—of his or her native language. This has been shown in several studies demonstrating that infants from birth to 4 months are able to discriminate the sounds of all languages, whereas at the end of the first year they no longer discriminate the foreign sounds but only those of their native language. This is illustrated by two studies from Werker and colleagues (1981, 1990), who showed this using English and Japanese in one group, and Hindi and Salish in another. All of the younger infants were able to distinguish all of the phonemes, but at 1 year of age the children raised in an English (Canadian) environment had lost their ability to distinguish the Japanese, and those raised in a Japanese environment had lost their ability to discriminate the English phonemes; the same results found for those raised in the Hindi or Salish contexts. The basic foundation of language—phonology—is formed by the end of the first year of life. This process of going from many to few (specialization) parallels the known neuroanatomical process of the reduction of dendrites in the developing CNS.

There are limited data as to the time dimensions of the critical/sensitive period(s) for semantics and syntax from a group of neurophysiological studies conducted by Neville et al (1992, 1997). The evidence is based on the achievement of a fixed (i.e., mature) neurophysiological response pattern. It is assumed that when the neurophysiological pattern becomes fixed, the plasticity is reduced or extinguished, and this decrease in neurophysiological plasticity is considered as the end of the critical/sensitive period. This is a tenuous set of assumptions based on few data and serves only as an approximate time frame for the end of the critical/sensitive period(s) of semantics and syntax.

Neville and colleagues recorded event-related brain potentials (ERPs) to words in hearing children, adolescents, and adults. They used two types of words. The first type is categorized as "open class" words, which consist of nouns, verbs, and adjectives that make reference to specific objects and events. These words are considered as semantic. The second group, "closed class" words, are articles, conjunctions, and auxiliaries. In the English language, these "closed class" words are part of the basis for the grammatical structures and may be thought of as an approximation of syntax.

Open class words (semantic representation) evoke in the normal adult a negative ERP occurring at 350 ms after the stimulus of the presentation of the word. The response is called the N350. The adult N350 is a large ERP found over the posterior brain regions of both hemispheres. The ERP to closed class words (syntax) in the normal adult occurs 280 ms after the stimulus. It is called the N280 and is found in the anterior temporal region of the left hemisphere.

Developmental studies found that, at 4 years of age, the youngest age tested, there were adult-like ERPs (N350) to open class/semantic stimuli. These N350 responses were robust and located at the posterior hemispheres. The N280 ERP, the response to the closed class/syntactic words, has a different ontogeny. The N280 does not achieve its mature, or final, configuration until 15 to 16 years of age. These data suggest that, based on ERPs, the critical period for semantic organization may occur before the fourth year of life, whereas the critical period of the assessed portion of syntax as evidenced by the closed class words may not occur until age 15 to 16.

Three aspects of aural language structure—phonology, semantics, and syntax—may have different time periods for optimal shaping of the CNS (i.e., the critical/sensitive period). The data suggest that the earliest specialization is phonological, and at the end of the first year the repertoire of phonemes that can be discriminated is limited. Phonological specialization comes before semantic specialization, and it is only suggested the critical period for phonological specialization may occur before 4 years of life. The last portion of the language structure to be set is syntactical, which will not be fixed until the late teens.

Although these time dimensions are approximations and certainly will be supplanted by more accurate and precise measures in the years to come, they do describe several important aspects of critical/sensitive period(s) for language acquisition. The concept/aspect is that different parts of language are governed by different biological constraints. Language does not happen in an instant, and full development is, in part, a time-dependent process.

A second concept is that different portions of the CNS, in a normal person, mediate different aspects of language. Up to now the investigation of language functions of the CNS have been explored based on the behavioral descriptions of language. These behavioral studies should produce information so that language could begin to be conceptualized based on the neurological mechanisms that underlie the process. The classification of language components based on the concepts of phonology, semantics, and syntax may or may not be the model for the neuroanatomical and physiological mechanisms that mediate the complex process called language.

Third, there appears to be a hierarchal sequencing of language development. The basis of language is the earliest restriction, in the case of an aural language, phonology. If there is little or no early phonological input, then the resultant semantic and syntactical portions cannot or do not develop optimally. The maturation of the semantics and syntax can be considered as being dependent upon early sensory input. It is well documented that either early auditory or visual linguistically organized sensory input will result in the development of sophisticated (i.e., functional) language. Much less is known about the informational characteristics of visual language. Are there any studies of the critical period of visual "phonology"? I know of no study that has looked to see if there would be the same restriction of aspects of the visual signs as in the phonological restrictions found at the end of 12 months in aural language. Are there signing characteristics that are found in American sign language (ASL) and not in Japanese sign language (JSL), and vice versa? Would an ASL child not be able to categorize the JSL at 12 months of age? It is known that if a child is given some sensory bases of language during the first 2 years of life, the semantic and syntactic aspects of that language develop into a useful and sophisticated communication system (effective language).

Neville and colleagues showed that deaf adults had normal N350, the ERP to open class/semantic words. However, these early-deafened adults, on the whole, lacked the N280 ERP to the closed set/syntactical words. It is unclear as to how, where, or even if syntax was being performed in these deaf adults. The few deaf adults who did have good grammatical judgment did have the expected left hemisphere N280s. This may indicate that the syntactical dimension may be an integral aspect of the left hemisphere specialization.

A recent series of retrospective studies by Yoshinaga-Itano (1998) examined the language outcomes in two groups of hearing-impaired children. The first group had a diagnosis and intervention with some form of language intervention before 6 months of age, and the second group had intervention and diagnosis when they were older than 6 months of age. Children who were diagnosed before 6 months of age had superior but not normal expressive and receptive language skills as compared with those who were diagnosed at the older age (i.e., greater than 6 months). These data strongly suggest that the earlier that language is instituted, the better the language outcome is. The researchers noted that the children with interventions before 6 months of age had a variety of interventions, including hearing aids, total communication, and/or sign language. Future types of linguistic input—auditory and/or visual—did not make any difference in the outcome. It has been documented that deaf children whose language is based solely on signs, on the average, will develop their first words somewhat sooner than the average hearing child exposed to normal aural language. Word order and morphology occur at the same age in both the sign-reared deaf and the auralreared hearing child. These observations show that linguistic input other than hearing will allow for the development of language. These observations suggest that the early formation of a language base is not dependent upon any particular sensory modality, but if the information is appropriately coded, language can develop from a sensory input that is not auditory. There is further evidence that this may occur from the education of the deaf blind in which the linguistic input is that of touch. Many of these patients, such as Laura Bridgman and Helen Keller, have had 1 or 2 years of aural language before they became deaf and blind. However, in both of these women and many others, the use of touch was sufficient to allow for complex linguistic development. That there are different sensory inputs such as auditory, visual, and tactile that can create bases for semantics and syntax indicates that language as described by semantics and syntax is not dependent upon any particular sensory form of input, but is dependent upon some form of linguistically organized sensory input.

There has been and continues to be controversy that a language initially learned with one sensory modality may not be transferable to one that is based on another sensory modality. More directly, a child learns sign language who is thought by many not to be able to develop optimal aural language. The establishment of language based on one sensory modality does appear to allow for the utilization of other sensory systems for the expression and reception of language, as has been shown. A study by Dee et al. (1982) performed some years ago with a group of, by today's standards, late-diagnosed deaf children in a total communication program found that 11 of the 12 developed speech and a few, for the time when this was done, excellent aural language. The 12th child had a tracheotomy and a cleft palate and was avocal.

CENTRAL NERVOUS SYSTEM PLASTICITY

The ability of various sensory inputs to allow for linguistic development suggests that the central nervous system is, to some extent, flexible (plastic) in regard to the formation and accomplishment of language. We have seen that interference with the establishment of the components of language, especially phonology in spoken language, results in substantial and apparently irrevocable linguistic deficits. Has nature created such a rigid structure that there is no redundancy or ability to compensate for deficiency? What is, if any, the ability of the CNS to compensate for defects? How plastic is the CNS for language?

There are numerous studies (Sharma et al. [1982] and van Melcher et al. [2000]) that demonstrate the plasticity of the CNS in response to a loss of a portion of the cochlear. The CNS has the ability to "cannibalize" the unused portion of the nervous system to respond to the remaining cochlear inputs. One of the more dramatic and recent has been a study in immature ferrets in which the visual systems were implanted into the auditory system in the newborn ferret. These ferrets grew up and saw with their auditory cortex.

The studies of Knudsen et al (1990, 2000) of sound localization in the barn owl are more pertinent to our interest in oral language. The researchers placed optical prisms on the eyes of newborn owls, resulting in a visual perversion of space, and found that the auditory CNS would reorganize itself to compensate for the distortion of vision. When they did the same to the mature owl, there was no compensation. The older owls' CNS could not adjust. These data would indicate that there was a very sharp critical period for the organization of the owls' auditory system, which was dependent upon visual and acoustic input. The study of the mature owl was carried a little further. The researchers fitted the mature owls with prisms that had very little distortion. The mature owl and its CNS compensated for this lesser perversion of visual space. The owls were then fitted with another prism that changed the visual field a little more and more progressively until the mature owls now had their visual field change so much that neither of the newborn nor the mature owl could compensate in one step. This is evidence of a quantitative change in plasticity in contrast to an all-or-none conceptualization of the process. It would appear that the owls' CNS maintained an attenuated plasticity.

King and colleagues (2000, 2001) have performed parallel studies, plugging one external auditory canal in a young ferret, causing a perversion of auditory space. They have found that the young ferret and its CNS would compensate for the "misinformation" about auditory space. This was evidenced by anatomical and physiological changes in the auditory CNS of the young ferret. When the same was done to the mature ferret, like the mature owl, they did not compensate. However, when given many trials, the adult ferrets recognized their CNS and were able to complete the task. Both of these studies suggest that the auditory CNS remains plastic throughout life, with a diminution of the amount of change it is able to accomplish as the organism ages.

The limits of plasticity or critical period for language are suggested but not directly addressed by the aforementioned studies. An indication of the human CNS's linguistic plasticity is to be found in the studies of young children who have had removal of either their right or left cerebral hemisphere. There are several indications for this procedure of which the hemispherectomies performed. One of these, Rasmussen's syndrome, is a unilateral progressive inflammatory disease characterized by normal development and late-onset seizures between 6 and 13 years of age. It allows for a more homogeneous population in which the possibility of the utilization of unusual brain areas for language may be studied. Several case studies have shown that right-handed children with left hemispherectomies recover and continue to develop linguistically. Their recovery recapitulates the usual development sequences: first, receptive skills are evident, followed in time by expressive language. Most of these children are able to carry on a conversation despite the loss of their left brain. It is not known whether the right brain was able to be the language center and is not used because of a preference for the left brain, or if the right brain is able to adapt so as to carry out language function. There are other conjectures, but the observations remain that the child's CNS is able to compensate for the severe injury. There is ability, given the circumstance of age and/or situation, for the CNS to adapt to change so that the person may be linguistically competent.

SUMMARY

The concept of a "critical/sensitive" period has undergone change in the past 2 decades. Today we must consider that the normal ontogeny for optimal language is dependent upon the institution of some form of linguistic input before the first 6 months of life. The CNS has an ability to adapt itself to the sensory changes—different or perverted inputs and/or intrinsic loss, such as with a hemispherectomy—that appears to be inversely proportional to age. This plastic tenet of the CNS does not entirely close down. The next frontier, the biology of language, will be to learn how to use this plastic ability to optimize the linguistic ability of each individual.

Suggested Readings

Birnholz JC, Benacerraf BR. The development of human fetal hearing. Science 1983;222(4623):516–518

Boatman D, Freeman J, Vining E, et al. Language recovery after left-hemispherectomy in children with late-onset seizures. Ann Neurol 1999;46(4):579–586

Bonvillian JD, Orlansky MD, Novack LL. Developmental milestones: sign language acquisition and motor development. Child Dev 1983;54(6):1435–1445

DeCasper AJ, Prescott PA. Human newborns' perception of male voices: preference, discrimination, and reinforcing value. Dev Psychobiol 1984;17(5):481–491

Dee A, Rapin I, Ruben RJ. Speech and language development in a parent-infant total communication program. Ann Otol Rhinol Laryngol Suppl 1982;97:62–72

Itard JMG. De l'education d'un homme sauvage ou des premiers développements physiques et moraux dy jeune sauvage de l'Aveyron. Paris: Goujon Fils; 1801

Jusczyk PW, Bertoncini J. Viewing the development of speech perception as an innately guided learning process. Lang Speech 1988;31(Pt 3):217–238

Knudsen EI, Knudsen PF. Sensitive and critical periods for visual calibration of sound localization by barn owls. J Neurosci 1990;10(1):222–232

Knudsen EI, Zheng W, DeBello WM. Traces of learning in the auditory localization pathway. Proc Natl Acad Sci USA 2000;97(22): 11815–11820

Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C. A precursor of language acquisition in young infants. Cognition 1988;29(2):143–178

Neville HJ, Coffey SA, Lawson DS, Fischer A, Emmorey K, Bellugi U. Neural systems mediating American sign language: effects of sensory experience and age of acquisition. Brain Lang 1997;57(3):285–308

Neville HJ, Mills DL, Lawson DS. Fractionating language: different neural subsystems with different sensitive periods. Cereb Cortex 1992;2(3):244–258

Petitto LA, Marentette PF. Babbling in the manual mode: evidence for the ontogeny of language. Science 1991;251(5):1493–1496

Querleu D, Renard X, Versyp F, Paris-Delrue L, Crepin G. Fetal hearing. Eur J Obstet Gynecol Reprod Biol 1988;28(3): 191–212

Querleu D, Renard X, Versyp F, Paris-Delrue L, Vervoort P. Intra-amniotic transmission of the human voice [in French]. Rev Fr Gynecol Obstet 1988;83(1):43–50

Rawlinson G, trans. The History of Herodotus. New York: Tudor; 1928

Ruben RJ. The ontogeny of human hearing. Acta Otolaryngol (Stockh) 1992;112(2):192–196

Sharma J, Angelucci A, Sur M. Induction of visual orientation modules in auditory cortex [see comments]. Nature 2000; 404:841–847

von Melchner L, Pallas SL, Sur M. Visual behavior mediated by retinal projections directed to the auditory pathway [see comments]. Nature 2000;404:871–876

Wardrop J. History of James Mitchell, a Boy Born Blind and Deaf with an Account of the Operation Performed for the Recovery of His Sight. London: John Murray; 1813

Werker JF, Gilbert JH, Humphrey K, Tees RC. Developmental aspects of cross-language speech perception. Child Dev 1981; 52(1):349–355

Werker JF, Tees RC. Influences on infant speech processing: toward a new synthesis. Annu Rev Psychol 1999;50:509–535

Yoshinaga-Itano C, Sedey AL, Coulter DK, Mehl AL. Language of early- and later-identified children with hearing loss. Pediatrics 1998;102(5):1161–1171

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Language development can progress
   - A. Only if there is hearing
   - B. If there is an organized sensory linguistic input
   - C. If the child is raised in isolation
2. The order of linguistic development is
   - A. Syntax, phonology, semantics
   - B. Phonology, syntax, semantics
   - C. Phonology, semantics, syntax
3. There is evidence that the sensitive/critical period for language acquisition
   - A. Is over by 18 months of age
   - B. Does not exist
   - C. Diminishes with age

Chapter 30
PRINCIPLES OF AUDIOMETRY

Jackson Roush and John Grose

Audiometry is concerned with the measurement of hearing and auditory function. Audiology, being broader in scope, encompasses not only the evaluation of hearing and its disorders, but also the myriad skills and techniques involved in hearing health care such as habilitation/rehabilitation of hearing loss, identification of hearing loss in newborn infants, the selection and fitting of hearing aids, evaluation and postsurgical management of patients with cochlear implants, assessment of vestibular function, and prevention of hearing loss through programs of hearing conservation.

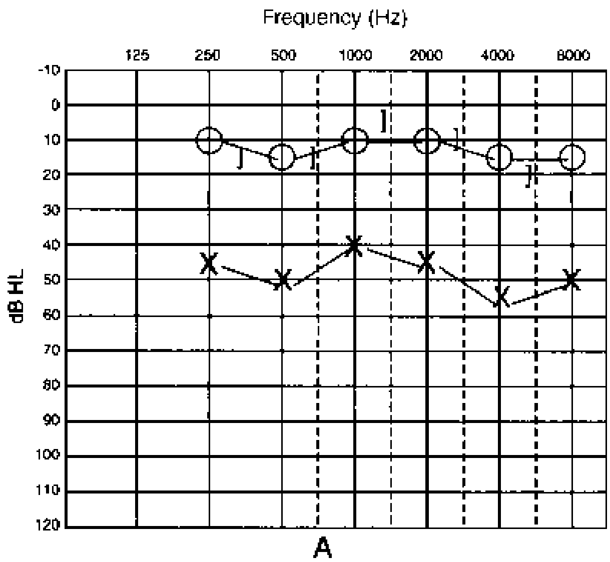
Audiometric assessment may be undertaken for a variety of reasons. In many cases it is aimed at quantifying the amount of usable hearing so as to determine whether a person's communicative abilities are at risk and, if so, whether they would benefit from a sensory device such as a hearing aid or cochlear implant. In other cases, assessing the functionality of the auditory system focuses more specifically on how intact that system is, or how well a particular component of that system is working. For example, the question of interest may be whether surgery has been successful in restoring mobility to the middle ear system. Because audiometry is multifaceted, encompassing procedures aimed at both an assessment of hearing and an assessment of auditory function, a useful framework in which to describe audiometry is to distinguish between subjective and objective procedures. Subjective procedures are based on behavioral responses from the listener and, in most cases, are the preferred methods to determine that a sound has indeed been heard. Objective procedures, on the other hand, do not rely on behavioral responses but instead use various physiological test procedures to measure the integrity or functionality of various components of the auditory system. The audiologist plays a critical role in the assessment process by selecting and applying test procedures appropriate to a given situation. Often this requires the application of a test battery, permitting a systematic "cross-check" of clinical findings based on a combination of behavioral and objective measures. This chapter will describe a general framework of audiometric indices, beginning with a discussion of behavioral assessment and then describing various objective procedures.

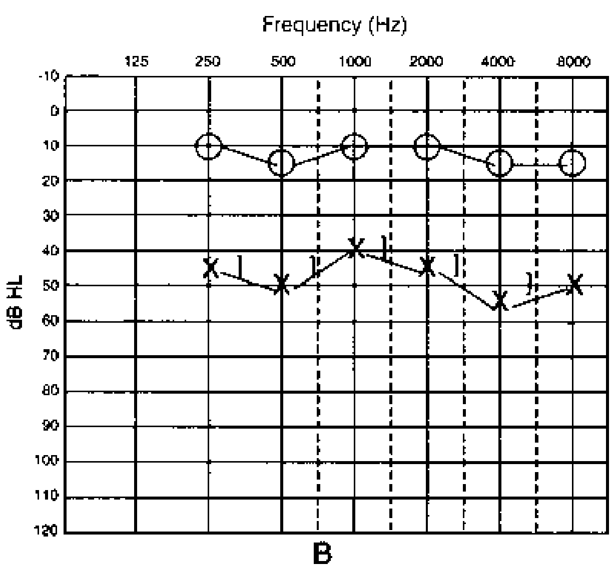
BEHAVIORAL ASSESSMENT OF HEARING

Pure-Tone Audiometry

Pure-tone audiometry is designed to measure thresholds of detection for pure-tone signals presented via air or bone conduction. Air conduction audiometry involves the presentation of test stimuli from an earphone, an insert receiver, or a loudspeaker, whereas bone conduction audiometry refers to the presentation of signals through a bone vibrator, usually placed behind the ear on the mastoid process. Testing is generally performed at octave intervals from 250 to 8000 Hz for air conduction and 250 to 4000 Hz for bone conduction. Detection thresholds are graphically displayed as an audiogram that plots threshold levels in decibels hearing level (dB HL) as a function of frequency using standard symbols (see **Fig. 30–1**). Degree of hearing loss can be summarized as the pure-tone average, based on the average of the air conduction thresholds at 500 Hz, 1000 Hz, and 2000 Hz. Descriptive terms commonly used to categorize hearing levels include normal (0–15 dB HL), borderline normal (16–25 dB HL), mild loss (26–45 dB HL), moderate loss (46–75 dB HL), severe loss (76–90 dB HL), and profound loss (> 90 + dB HL). Borderline categories of hearing impairment are often described with a combination of terms, such as mild to moderate, as are sloping configurations due to differences in hearing sensitivity across the frequency range.

Modality	Ear: LEFT	Ear: RIGHT
Air conduction		
Unmasked	X	O
Masked	□	Δ
Bone conduction		
Unmasked	>	<
Masked	]	[





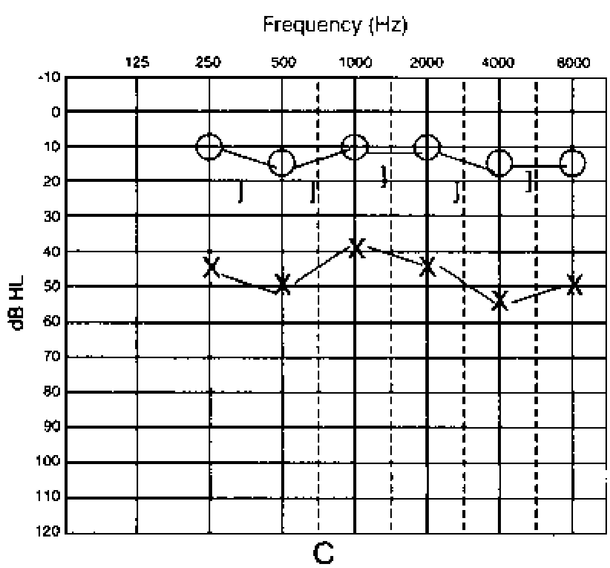


Figure 30–1 *Audiograms illustrating **(A)** a moderate unilateral conductive hearing loss, **(B)** a moderate unilateral sensorineural hearing loss, and **(C)** a moderate unilateral mixed hearing loss. In all three examples the right ear is within normal limits. (Adapted from Silman S, Silverman C. Auditory Diagnosis. San Diego: Academic Press, 1991.)*

Methods used to obtain the behavioral pure-tone audiogram vary depending on the patient's age and developmental status. Adults and children at a developmental age of 5+ years raise their hand or press a response button when the tone is heard; toddlers require a play audiometry task, and infants from 6 to ~30 months require an operant conditioning procedure known as visual reinforcement audiometry (VRA). For infants at a developmental age less than 6 months, behavioral test procedures are not considered valid; thus hearing status must be inferred from objective measures described later in this chapter.

When pure-tone air conduction thresholds are abnormally elevated, bone conduction thresholds are obtained to differentiate between sound transmission problems lateral to the inner ear and dysfunctions of the auditory system in or medial to the inner ear. Conductive hearing loss is characterized by normal or near normal bone conduction thresholds in the presence of elevated air conduction thresholds (**Fig. 30–1A**). When air and bone conduction thresholds are equally elevated, the loss is described as sensorineural (**Fig. 30–1B**). A mixed hearing loss (**Fig. 30–1C**) is characterized by abnormal responses to both air conduction and bone conduction signals, with air conduction thresholds being poorer than bone conduction thresholds. It should be noted that some middle ear pathologies, in addition to resulting in conductive hearing losses, may show elevated bone conduction responses at or around 2000 Hz due to loss of the normal middle ear participation in the bone conduction response. This pattern of results, described as Carhart's notch, can be diagnostically relevant in the context of reviewing patient candidacy for middle ear surgery. In such cases a more accurate estimate of sensory loss can be achieved using a forehead placement of the bone vibrator, because the effects of middle ear pathology on the bone conduction response are diminished using frontal placement. It should be noted that adjustments in calibration are needed for forehead placement.

Pure-tone audiometry may be complicated by transmission of test signals across the skull to the nontest ear. The audiologist isolates the test ear by applying masking noise to the nontest ear. Because diagnosis and treatment are often influenced by the audiometric findings, accurate use of masking noise is imperative.

Pediatric Modifications in Pure-Tone Audiometry

An operant conditioning procedure, based on reinforcement of head-turn responses, is useful for the behavioral assessment of hearing in infants beginning at a developmental level of ~6 months. This technique, called visual reinforcement audiometry, involves placing the infant in a calibrated sound room while test signals (frequency modulated pure tones) are presented from a loudspeaker. An assistant is seated across from the child to maintain attention away from the loudspeaker and a reinforcer (illuminated mechanical toy in a smoked-glass enclosure). Once the infant is conditioned, stimuli are presented at lower levels to determine an approximate threshold. Testing can be performed with stimuli presented from a loudspeaker or via insert receivers. When test signals are delivered from a loudspeaker, the responses are considered an indication of "better ear" sensitivity for each test frequency. Responses obtained via insert receivers permit evaluation of each ear separately. The VRA technique also can be used to obtain bone conduction thresholds.

When the child reaches a developmental level of ~30 months, it is generally feasible to conduct manual pure-tone "play" audiometry. The response task for older children (5+ years) involves raising a hand or pressing a response button when the tone is heard; younger children (30 months–4 years) respond by engaging in some age-appropriate play activity, such as dropping a block in a container or placing a peg in a board. Play audiometry is usually performed under earphones (or with insert receivers) so that each ear can be evaluated separately. Bone conduction thresholds are obtained in the same manner.

Speech Audiometry

Routine speech audiometry consists of speech threshold measures and speech recognition measures. The speech threshold test, also known as the speech reception threshold (SRT), is performed using two-syllable spondee words that have equal emphasis on both syllables (e.g., *baseball*). SRTs are obtained in a manner similar to pure-tone threshold assessment. Speech recognition measures, more commonly referred to as speech discrimination tests, are typically performed using lists of prerecorded single-syllable words presented at a fixed level above threshold. To illustrate, a patient with a mild hearing loss and 35 dB HL pure-tone average would be expected to have an SRT of ~35 dB HL because the SRT is normally in general agreement with the pure-tone average (average of thresholds at 500, 1000, and 2000 Hz). The speech discrimination test would then be administered at a level ~40 dB above the SRT (i.e., at a 40 dB "sensation level"), which, in this example, would be at 75 dB HL. The SRT serves as a cross-check on the validity

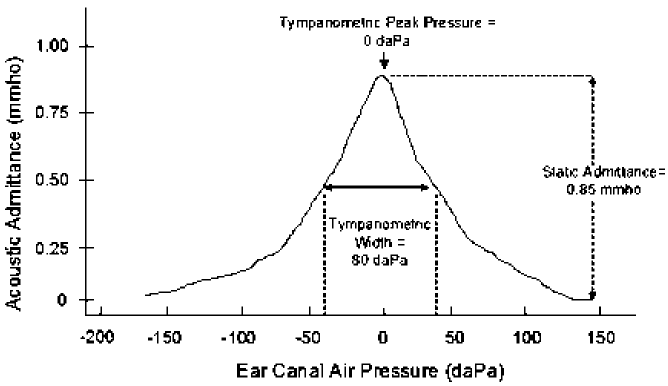


Figure 30–2 *Measures derived from the standard (226 Hz) tympanogram.*

of the pure-tone responses as well as a reference level for speech discrimination testing. The speech discrimination test provides a qualitative measure of understanding for speech in quiet, delivered at or near a comfortable level above threshold; the score is expressed in percent correct for each ear. Taken together, these measures are useful for diagnostic purposes and in the planning of rehabilitation (e.g., hearing aid use). In addition to single words presented in a quiet background, many other speech recognition measures have been developed for diagnostic and rehabilitative applications.

OBJECTIVE MEASURES OF AUDITORY FUNCTION

Immittance Audiometry

The assessment of middle ear function requires a battery of procedures that measure the ease with which a sound is transmitted through the conductive mechanism. This battery, which includes tympanometry and a series of related measures, provides information concerning the mobility of the tympanic membrane and middle ear system, middle ear pressure, and equivalent volume estimates. Each of these indices may be associated with various middle ear abnormalities.

Tympanometry begins with a sealed probe tip in the ear canal. The probe assembly provides both a sound source and a measuring microphone. As the air pressure in the ear canal is varied by the tympanometer, the associated movements of the eardrum and middle ear system are detected as amplitude and phase changes relative to a constant low-frequency probe tone (200/226 Hz) measured from a microphone in the probe assembly. These changes are due to alterations in the physical impedance to the sound at the plane of the tympanic membrane, as positive and negative pressure causes changes in the stiffness of the tympanic membrane. In essence, they provide a measure of the ease with which the probe tone is admitted through the tympanic membrane and middle ear system. When plotted as a function of the relative pressure in the sealed ear canal, the admittance changes result in a graphic representation known as the tympanogram, which may be thought of as an "objective" form of pneumatic otoscopy. A complete acoustic immittance battery includes measures of tympanometric peak pressure, tympanometric shape (width), static admittance, and estimates of equivalent ear canal volume, all of which are derived from the tympanogram (**Fig. 30–2**).

Tympanometric peak pressure, the pressure in the external canal when static admittance is highest, is measured in decapascals (daPa) and provides an estimate of middle ear pressure. As illustrated in **Fig. 30–3A**, a normal tympanogram will have a peak that approximates ambient air pressure (−50−+150 daPa). Tympanometry performed on an ear with middle ear effusion typically results in a "flat" tympanogram (immobile or noncompliant; **Fig. 30–3B**). The tympanogram shown in **Fig. 30–3C** has a pressure peak at approximately −240 daPa, meaning that middle ear admittance was maximum when the tympanometer created negative pressure (−240 daPa) in the ear canal (i.e., equal pressure on either side of the tympanic membrane). In this case, it can be assumed that the middle ear is characterized by negative pressure, presumably due to insufficient aeration of the middle ear space. It should be noted, however, that changes in middle ear pressure are caused by a complex interaction of many factors. Consequently, tympanometric peak pressure is of limited value in the detection of middle ear dysfunction.

Static measures of tympanic membrane mobility also can be obtained. Static admittance, which is measured in acoustic millimhos (mmho), is calculated by measuring the height of the tympanometric peak and comparing it to one of its tail values. Patients with otosclerosis or ossicular fixation usually show reduced middle ear mobility, evidenced by low static admittance. The resulting tympanogram often has a shallow peak with normal middle ear pressure. Static admittance measures are useful for detecting middle ear effusion. For children 3 to

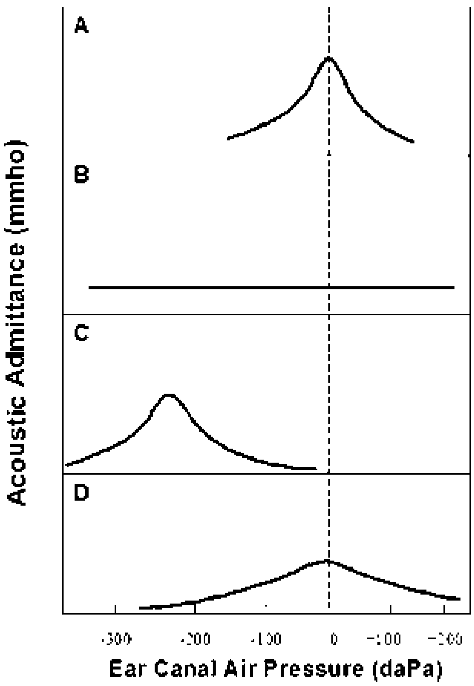


Figure 30–3 *Tympanometric patterns commonly associated with **(A)** normal middle ear function, **(B)** middle ear effusion, **(C)** negative tympanometric peak (middle ear) pressure, and **(D)** an abnormally wide configuration that may or may not be associated with middle ear disease.*

10 years of age with no history of chronic or recurrent otitis media with effusion (OME), mean values are typically 0.5 mmho with a 90% range of 0.3 to 1.0 mmho. Infants and toddlers, especially those with significant OME histories, usually demonstrate lower peak static admittance, even when middle ear effusion is not present. Because middle ear effusion reduces the mobility of the middle ear mechanism, a tympanogram associated with OME is typically characterized by low static admittance (0–0.2 mmho). The flat tympanogram in **Fig. 30–3B** is typical of that seen in a child with OME.

The shape of the tympanogram can be altered by middle ear effusion even when a pressure peak is identifiable. For most screening instruments, tympanometric width is reported as the pressure interval (in daPa) corresponding with a 50% reduction in static admittance (see **Fig. 30–2**). Tympanometric width has proven to be useful in the detection of middle ear effusion. The tympanogram is usually flat when middle ear effusion is present, but wide tympanograms (e.g., >200 daPa) may also be associated with middle ear effusion. The tympanogram in **Fig. 30–3D** is abnormally wide. Such a tympanogram may occur when middle ear effusion occupies a portion of the middle ear space while still allowing some aeration of the middle ear.

Estimates of equivalent ear canal volume are useful in identifying a perforation of the tympanic membrane or confirming tympanostomy tube patency. In either case the middle ear space is added to the ear canal space, resulting in an abnormally large equivalent volume (e.g., 2.5 cc). Estimates of equivalent ear canal volume are especially useful in the interpretation of "flat" tympanograms. Flat tympanograms may be due to (1) perforation of the tympanic membrane, a condition that would result in an abnormally large equivalent canal volume; (2) OME, which would result in a normal equivalent canal volume; or (3) occlusion of the probe tip (cerumen or canal obstruction), which would show abnormally small equivalent volume.

The acoustic reflex occurs when a high-intensity sound causes contraction of the stapedius muscle, momentarily restricting the transmission of low-frequency sound through the middle ear due to the stiffening in the middle ear. When used clinically, the acoustic reflex is elicited by a pure tone delivered via the tympanometer probe assembly. Instruments designed for middle ear assessment typically deliver ipsilateral or contralateral activating stimuli. The acoustic reflex is generally absent in an ear with middle ear dysfunction because the conductive hearing loss reduces the level of the acoustic stimulus reaching the inner ear and/or because the stiffness caused by the middle ear disorder prevents its measurement. Inclusion of the acoustic reflex in acoustic immittance screening batteries has been shown by some studies to improve the identification of ears with otitis media, but this often occurs at the expense of lower specificity (i.e., a higher number of false-positive referrals).

The acoustic reflex has many clinical applications. In the 1970s and 1980s, the acoustic reflex was used to screen for retrocochlear disease. Since then, it has been replaced by tests with better predictive value. It is still useful, however, as a test of cranial nerve (CN) VII and VIII function. It is also useful in the context of screening for middle ear effusion when the tympanogram is ambiguous. A flat tympanogram is expected when middle ear effusion is present, but as noted earlier, middle ear effusion can be present even when the tympanogram has an identifiable peak, especially in the case of wide tympanograms (**Fig. 30–3D**). Because the acoustic reflex will almost always be absent when middle ear fluid is present, the acoustic reflex can assist in differentiating wide tympanograms associated with effusion from those without effusion.

Otoacoustic Emissions

Basis of OAEs

The inner ear, or cochlea, is characterized by an active mechanism that has the effect of amplifying low-level input (see Chapter 25). The "motor" for this active mechanism is associated with the outer hair cells, which vary in length according to the surrounding electrical potentials. The rapid oscillations in outer hair cell length "pump" energy back into the vibrating basilar membrane by increasing the amplitude of vibration in localized regions. An epiphenomenon of this active feedback mechanism is that the vibrations introduced into the basilar membrane motion by the outer hair cells can be transmitted back out through the middle ear system in a reversal to the usual direction of sound flow. The tympanic membrane, which is usually thought of as analogous to the diaphragm of a microphone, now acts analogously to the cone of a loudspeaker. A sensitive microphone sealed into the exterior auditory meatus can detect the low-level sounds generated by the outer hair cells and subsequently the vibrations of the tympanic membrane. These low-level sounds are known as otoacoustic emissions (OAEs).

Two important points deserve note. First, the active mechanism of the inner ear is associated with normal cochlear function, and therefore any impairment of cochlear function usually results in a failure of the active mechanism. Thus OAEs are a signature of normal cochlear function and generally are not present in cochlear-impaired ears. Second, the detection of OAEs in the external auditory meatus requires not only that the inner ear is functioning normally but also that the transmission pathway back through the middle ear is functioning adequately. Thus middle ear dysfunction is likely to reduce or obliterate successful recording of OAEs. Taken together, these two points indicate that successful recording of OAEs generally points to both normal cochlear function and normal middle ear function, whereas absence of OAEs could be due either to cochlear impairment or to middle ear dysfunction. Additional differential testing is required in the latter case.

Types of OAEs

Although the mechanism of OAE generation is probably singular, different categorizations of OAEs have emerged based on the stimulus characteristics required to elicit them. The two most common OAE categories encountered in clinical practice are transient evoked otoacoustic emissions (TEOAEs) and distortion product otoacoustic emissions (DPOAEs).

As the name implies, TEOAEs are evoked with a transient acoustic stimulus, such as a click. An example of a normal TEOAE recording from a 1-day-old infant is shown in the upper half of **Fig. 30–4**. (For comparison, a recording from a newborn with an absent TEOAE is shown in the lower half.) In each half of the figure, panel A shows the time waveform of the click in the ear canal, and panel B (top line) shows the energy distribution of the click in the frequency domain. Following the click presentation, the OAE emerges after a short delay. In this context, the TEOAEs are sometimes referred to as "echoes" (or "Kemp echoes," after the physicist who discovered them). Panel C shows the time waveform of the TEOAE. The first few milliseconds of the trace are removed to avoid contamination by the stimulus artifact. Although not evident in the truncated response shown in **Fig. 30–4**, the higher frequencies of the emission emerge first (they have shorter latencies), and the lower frequencies emerge later. This reflects the travel time along the basilar membrane from the basal high-frequency regions to the apical low-frequency regions. Panel D shows the normalized spectrum of the response. Because of the transfer function of the middle ear, the response is usually dominated by the midfrequencies. The signal-to-noise ratio in different frequency bands is summarized in panel E, which allows decisions to be made concerning the presence or absence of a response.

The other common OAE category encountered in clinical practice is DPOAEs. To evoke these, two pure tones are presented to the ear whose relative levels and frequency relations are optimized to generate an intermodulation distortion product. Because of the inherent nonlinearity of normal cochlear function, intermodulation between two primary tones (termed F1 and F2) can generate distortion products that are physically present within the cochlea. The predominant distortion product has the frequency 2F1-F2, and reversed transmission of this tonal distortion product renders it recordable in the ear canal as a DPOAE. Because of the fixed frequency relation between the 2F1-F2 distortion product and the eliciting pair of primary tones, the frequency of the DPOAE can be varied by shifting the frequencies of the primary tones. A systematic probing of different frequency regions using this technique allows for the generation of a distortion product-gram (DPgram), a plot of DPOAE level as a function of primary tone frequency. An example of a DPgram from an adult with high-frequency hearing loss is shown in **Fig. 30–5**. The connected open circles refer to the levels of the DPOAEs, and the lower hatched area denotes the noise floor levels. Most forms of the DPgram depict both the DPOAE levels and the noise floor because the decision as to whether a valid DPOAE is present includes a consideration of the relative signal-to-noise ratio.

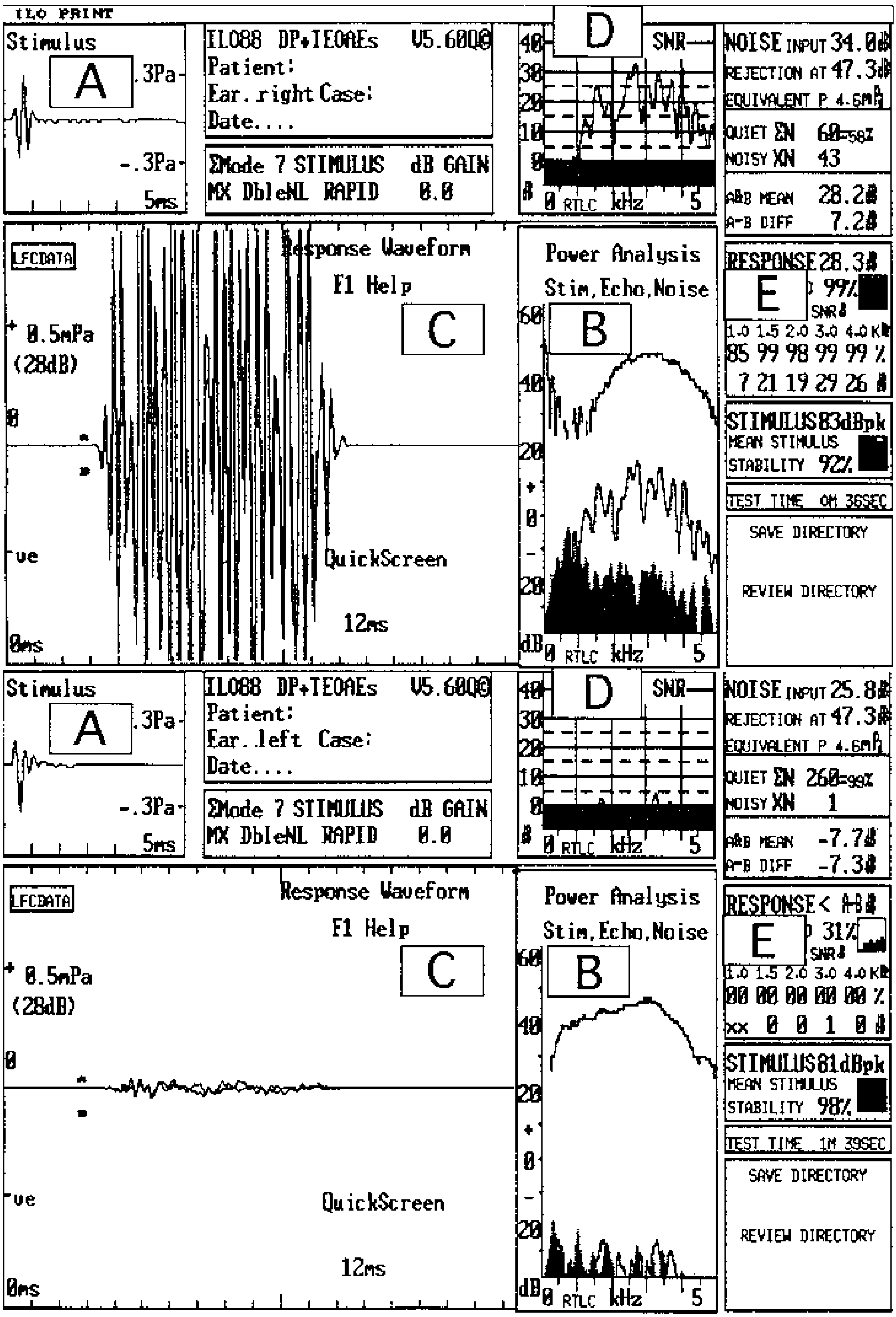


Figure 30–4 *Sample recordings from two newborn infants, one with present transiently evoked otoacoustic emissions (TEOAEs; upper half of figure) and the other with absent TEOAEs (lower half of figure).* ***(A)*** *Time waveform of click stimulus.* ***(B)*** *Spectrum of click stimulus (upper solid line), TEOAE (lower solid line), and noise floor (shaded area).* ***(C)*** *Time waveform of TEOAE.* ***(D)*** *Normalized spectrum of TEOAE.* ***(E)*** *Signal-to-noise ratios and response reproducibility within defined frequency bands of the response.*

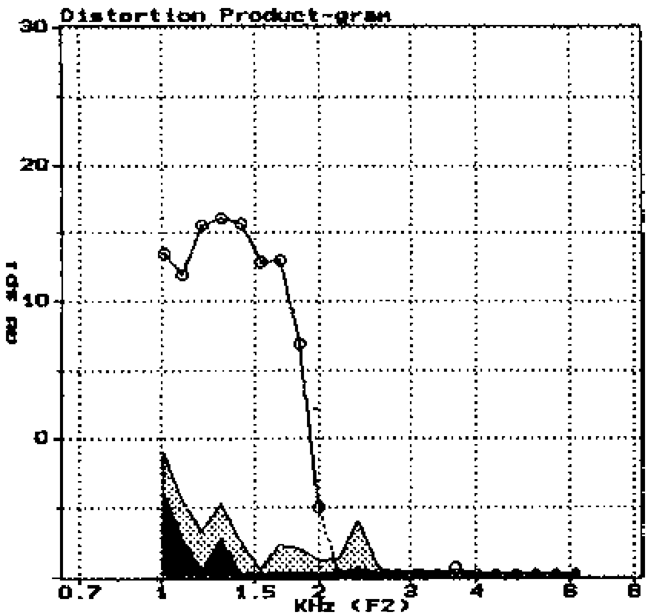


Figure 30–5 *Distortion product-gram from an ear with a high-frequency hearing loss (>1.5 kHz). The levels of the distortion product otoacoustic emissions (open circles) are plotted as a function of the frequency of the upper primary tone (F2). Lower hatched area indicates noise floor.*

Clinical Applications of OAEs

The main clinical application of OAEs is as a screening tool for identifying sensory dysfunction. Because OAEs are associated with normal cochlear function and are present even at birth, ears of any age with cochlear hearing losses greater than ~30 dB HL do not typically exhibit OAEs. Therefore, failure to elicit an OAE after ruling out middle ear dysfunction suggests the presence of a sensory loss. Using OAEs as a screening tool is popular, particularly for neonatal hearing screening programs, because the test is quick and noninvasive. Many clinics are moving toward routinely using OAE testing in protocols for monitoring cochlear health. Most agents and conditions that are detrimental to hearing (e.g., noise exposure, ototoxic medications, and aging) have a primary effect at the level of the cochlea, so it is appropriate to monitor for their effects using OAE testing. The second important function served by OAE testing is confirmation of audiometric configuration. Audiograms that show regions of both normal and impaired hearing (e.g., high-frequency hearing losses) should be mirrored by the DPgram or spectral content of the TEOAE (see **Figs. 30–4** and **30–5**). That is, OAE energy should be evident in regions where the audiogram indicates normal hearing and should be absent in regions where the audiogram shows regions of hearing loss in excess of ~30 dB HL. Thus OAE testing provides useful confirmation of audiometric configuration in difficult-to-test patients, patients who cannot provide voluntary responses (e.g., very young or sedated patients), and patients whose voluntary responses are suspect due to compromised developmental function. It must be emphasized that once hearing losses are mild to moderate, OAE testing cannot provide an indication of severity of cochlear loss; a hearing loss could be moderate or profound, yet the OAE test will simply indicate an absence of response.

Advantages and Disadvantages of OAEs

In addition to being an objective measure of cochlear function, an advantage of OAEs is that their measurement is quick, noninvasive, and not generally subject to the state of the patient in terms of sleep, sedation, or alertness. They also provide a wide-band test of peripheral auditory function, unlike some evoked potential measures (see Evoked Potential Audiometry).

A main disadvantage of OAEs is that they do not provide a robust indication of degree of hearing loss. That is, once hearing losses exceed the mild range, OAEs cannot provide an indication of severity of cochlear loss. Because OAEs reflect outer hair cell function, it is possible to measure robust OAEs in the presence of a hearing loss that has a neural basis. This general condition is known as auditory neuropathy. Another disadvantage of OAEs is that their recording is dependent on middle ear status. Thus failure to record an OAE requires further differential testing, including an acoustic immittance battery, for interpretation. Finally, the measurement of OAEs is highly subject to ambient noise levels, generated by the patient (e.g., respiratory noise) or by external conditions (e.g., ventilators).

Evoked Potential Audiometry

Basis of Evoked Potentials

The process of transducing the vibrations of sound into discrete neural impulses in the cochlea involves electrochemical mechanisms. Thereafter, the flow of auditory information toward the cortex consists of neural code, and this neural activity, like all neural activity, is essentially electrical in nature. The electrical potentials generated by cochlear and neural structures are volume-conducted to the surface of the skull/scalp and, under certain conditions, can be detected using contact electrodes. The electrodes themselves cannot distinguish between electrical activity associated specifically with audition and electrical activity associated with other neural or myogenic systems (although the electrodes can be positioned on the skull to optimize the recording of activity from generators more specific to the auditory pathway). However, the minute electrical activity associated specifically with audition can be extracted from the ubiquitous neural "noise" by a variety of signal processing techniques, including synchronizing an averaging process to the auditory stimulus. That is, if the surface-recorded (far field) electrical activity is sampled only at the precise moment that an auditory stimulus is delivered, and if this sampling is repeated many times and summed, then only the stimulus-related activity will add together, whereas the remaining activity (which is random with respect to the auditory stimulus) will tend to cancel out. Evoked potential audiometry is accomplished using a signal-averaging computer to extract sound-related sensory and neural activity from the milieu of ongoing neural and myogenic activity.

Types of Evoked Potentials

The evoked potentials associated with audition can be subcategorized either on the basis of their temporal relation to the stimulus or on the basis of the nature of the stimulus that elicits them. The time of occurrence of an evoked potential relative to the onset of the stimulus is known as its latency.

The evoked potentials associated with cochlear activity are the cochlear microphonic (CM), the summating potential (SP), and the whole-nerve, or compound, action potential (WNAP/CAP). The CM is an alternating current (AC) potential that represents summed outer hair cell responses to sound; the SP is a direct current (DC) potential that arises because of the nonlinearity of cochlear function; and the WNAP represents the summed, synchronous "firing" of primary auditory neurons in response to the onset of stimulation. Potentials like the WNAP that depend on synchronous neural firings for their detection are best elicited with transient stimuli such as clicks and tone bursts. The clinical measurement of gross cochlear potentials is known as electrocochleography (ECochG). The recording requires an electrode in the vicinity of the cochlea, placed either extratympanically or transtympanically.

The synchronous firing of primary auditory neurons also can be recorded as an evoked potential using electrodes on the outer surface of the scalp. If the recording window is extended out to ~10 to 20 msec poststimulus, then evoked potentials associated with other auditory nuclei in the brainstem can also be recorded. Because these short-latency evoked potentials reflect activity through the auditory periphery and the brainstem, they are known collectively as the auditory brainstem response (ABR). The ABR is a sequence of several vertex-positive potentials, or waves, of which the first five (labeled with roman numerals I–V) are the most commonly examined. Responses from a 5-year-old child with one normal ear and one nonfunctional ear are shown in **Fig. 30–6**.

If the poststimulus recording window is extended still further, evoked potentials associated with cortical activity can be recorded. These include the middle latency responses (MLRs), with latencies out to ~80 msec, and the late potentials with latencies out to almost half a second. A variety of specific responses falls under the latter category, such as the P300 and mismatch negativity (MMN) responses. Other types of evoked potentials are measured using stimulation at special rates (e.g., the 40 Hz response that enhances evoked potential recording)

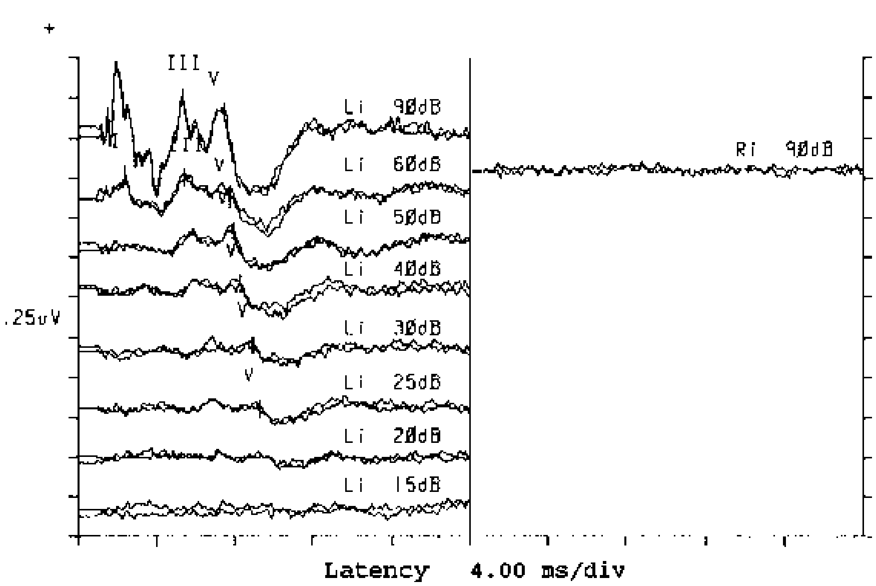


Figure 30–6 *Auditory brainstem response recordings from a 5-year-old child with one normal ear and one unresponsive ear.*

or to continuous tones (frequency following response). This discussion will confine itself to ECochG and the ABR.

Clinical Applications of Evoked Potentials

Some clinics make use of ECochG recordings to aid in the diagnosis of diseases associated with endolymphatic hydrops such as Meniere's disease. In these conditions, there is evidence that the ratio of the amplitude of the SP to the amplitude of the WNAP is enhanced. The procedure also may be used as a cross-check in the evaluation of infants and young children suspected of CN VIII dysfunction or auditory neuropathy.

The ABR has two major clinical applications: threshold estimation and differential diagnosis of retrocochlear dysfunction. For threshold estimation, the component of interest is wave V. As the intensity of the stimulus is lowered, the amplitude of this component is reduced, and its latency increases. However, it remains detectable down to stimulus levels that are within 10 to 15 dB of behavioral threshold depending on the type of stimulus being used. The most common stimulus used in the ABR is a click. This elicits highly synchronous neural responses, but the response is dominated by the midfrequency region. Tone burst stimuli can provide more frequency-specific threshold information, but there is a trade-off between the frequency specificity of the stimulus and its ability to evoke synchronous neural firing (and therefore a measurable response). In particular, latency-intensity response functions to low-frequency tone bursts typically are less clearly defined. Stimuli also can be delivered through a bone conductor and a comparison between the air-conducted response and the bone-conducted response can be helpful in determining the type of hearing loss. Threshold measures using the ABR allow an estimation of audiometric sensitivity to be formed that does not rely on the active cooperation of the patient.

The ABR traditionally has played a role in site of lesion testing. A retrocochlear lesion, such as an acoustic neuroma, can affect the morphology of the response waveform. Because the latencies of the ABR peaks for a given stimulus and intensity are highly repeatable, and similar across individuals within an age group, departure from normative values can contribute to a diagnosis of retrocochlear dysfunction. The increasingly widespread use of imaging techniques, such as magnetic resonance imaging (MRI), has impacted the role of ABR in site of lesion testing. However, it is worthwhile remembering that, whereas MRIs assess structure, the ABR is inherently a test of function. Therefore, its routine use in the comprehensive workup beyond MRI is justified.

Advantages and Disadvantages of Evoked Potentials

As with OAEs, the ABR is not sensitive to the state of the patient in terms of sleep or sedation. Measurable ABRs can be obtained even from premature infants as young as ~30 weeks gestation. Thus the ABR can be used to assess objectively auditory function in a wide variety of patient groups.

A disadvantage of the ABR is that it is inherently a test of synchronous neural response and therefore not an actual test of hearing. Its recording requires a quiet and still patient, and, for young children, this may require the use of conscious sedation. This often compounds the logistics and cost of administering the test because of patient-monitoring requirements. Whereas the clearest response is usually measured using the ubiquitous click stimulus, the response to this stimulus reflects only a portion of the cochlea. Finally, the ABR can be subject to electrical interference from other external sources.

SUMMARY

This chapter has provided an overview of the test battery commonly used to evaluate hearing sensitivity and auditory function. The information generated by these measurement tools is substantial, but like all assessment procedures, they must be applied appropriately and interpreted carefully. Moreover, obtaining these measurements is often only the first step. Deciding on an optimal course of action such as hearing aid selection or cochlear implantation and following through with appropriate rehabilitation strategies are all necessary components of the complete audiological practice. Optimal patient management requires mutual understanding and effective collaboration between the audiologist, the otolaryngologist, and related health care professionals.

Suggested Readings

Hall JW III. Handbook of Auditory Evoked Responses. Boston: Allyn & Bacon; 1992

Hall JW III. Handbook of Otoacoustic Emissions. San Diego: Singular Publishing Group; 2000

Hood L. Clinical Applications of the Auditory Brainstem Response. San Diego: Singular Publishing Group; 1998

Katz J, ed. Handbook of Clinical Audiology. Philadelphia: Lippincott Williams & Wilkins; 2002

Musiek FE, Rintelmann WF, eds. Contemporary Perspectives in Hearing Assessment. Boston: Allyn & Bacon; 1999

Robinette MS, Glattke TJ, eds. Otoacoustic Emissions: Clinical Applications. New York: Thieme Medical Publishers; 2002

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Behavioral assessment of hearing in an infant should not be attempted until the child has reached a developmental age of approximately
   A. 1 month
   B. 6 months
   C. 12 months
   D. 2 years
2. The appropriate behavioral assessment procedure for a typically developing 12-month-old infant is
   A. Visual reinforcement audiometry
   B. Play audiometry
   C. Immittance audiometry
   D. Otoacoustic emissions
3. The appropriate behavioral assessment procedure for a typically developing 3-year-old child is
   A. Visual reinforcement audiometry
   B. Play audiometry
   C. Immittance audiometry
   D. Otoacoustic emissions
4. Middle ear disease, in addition to creating a conductive hearing loss, may also reveal elevated pure-tone bone conduction responses at or around ____ Hz, due to loss of the normal middle ear participation in the bone conduction response.
   A. 250 Hz
   B. 750 Hz
   C. 2000 Hz
   D. 6000 Hz
5. The bone conduction shift described in question 4 is referred to as
   A. Mixed hearing loss
   B. Interaural attenuation
   C. Kemp's notch
   D. Carhart's notch
6. A flat tympanogram would not be caused by
   A. A patent tympanostomy tube
   B. Middle ear effusion
   C. Occlusion of the probe tip
   D. Endolymphatic hydrops
7. Otoacoustic emissions have the advantage of
   A. Providing an accurate indication of degree of hearing loss
   B. Providing a wide-frequency band test of cochlear function
   C. Not being affected by middle ear status
   D. Being measurable in noisy environments
8. A disadvantage of auditory brainstem response (ABR) testing is
   A. The response is obscured by sedation.
   B. The response reflects only synchronous neural activity.
   C. The ABR cannot be measured in infants until they are ~6 months of age.
   D. The test requires the use of transtympanic electrodes.

Chapter 31

HEARING AIDS, BONE-ANCHORED HEARING AIDS, AND COCHLEAR IMPLANTS

Adrien A. Eshraghi, Susan B. Waltzman, Joseph G. Feghali, Thomas R. Van De Water, and Noel L. Cohen

HEARING AIDS

Technology

The basic role of a hearing aid is to amplify auditory stimuli. A transducer, the microphone, which converts the signal from mechanical to electrical energy, picks up the incoming auditory signal. The electrical current is then amplified and transmitted to a receiver, which reconverts the electrical current into acoustic stimuli. The devices are battery-powered. There are many other elements to hearing aids that allow both the dispenser and the patient to adjust the device according to specific needs. Patient adjustment options include volume controls, on/off switches, and telephone and noise suppression switches, to name a few. The dispenser can alter the frequency response characteristics and program other acoustic variables to customize the device to the specific auditory needs of the patient. Currently, the most commonly used hearing aids are either in the ear (ITE), in the canal (ITC), completely in the canal (CIC), or behind the ear (BTE). Although the BTE was for a long time the most widely prescribed model, the other smaller and more cosmetically appealing types are gaining increasing popularity as they become more sophisticated and can provide benefit to a wider population.

A primary determinant of hearing aid selection relates to the circuitry that would best serve the needs of the patient. The simplest circuit, a linear design, amplifies the incoming signal in a predetermined manner regardless of the input level. A simple compression circuit reduces the loudness of selected sounds above a certain predetermined level that serves to minimize distortion. More developed compression circuits can more aggressively select the incoming signals to be modified. Although hearing aids have employed analog speech-processing circuits, more recent technology allows for analog processing with digital programming capability; that is, the adjustments are digitally driven, but the speech processing remains an analog function. Fully digital hearing aids that do employ digital signal processing, however, are becoming more accessible. The flexibility of the circuitry permits the device to store several different programs that can be used in different listening conditions. Recently developed multimicrophone, multimemory digital hearing aids deliver improved hearing in suboptimal listening situations.

Medical Evaluation

When possible, diseases of the external, middle, and inner ears, as well as diseases; of the auditory nerve and central auditory pathways, should be treated appropriately prior to receiving medical clearance for the use of a hearing aid.

Several medical conditions require additional testing to rule out significant disease; for example, acoustic neuroma in patients with significantly asymmetric and unexplained senorineural hearing loss, active autoimmune inner ear disorders, and Meniere's disease. However, these conditions are not necessarily absolute contraindications to the use of hearing aids. For example, it is permissible for a patient with a known acoustic neuroma to use a hearing aid if there are no treatments planned for the foreseeable future.

Otorrhea and chronic suppurative otitis media can interfere with a patient's ability to wear a hearing aid. Such conditions should be treated. Occasionally, these conditions persist, making it impossible for some patients to wear a hearing aid, and require the use of alternative treatments (e.g., bone-anchored hearing aids or other amplification modalities).

Hearing aids typically are dispensed by otolaryngologists, audiologists, and other hearing aid specialists. Some clinical situations require the direct involvement of an otolaryngologist or an otologist in the fitting process. This direct physician involvement is of particular importance in patients with open mastoid cavities, tympanic membrane perforations, and severe exostosis. In all these situations, ear mold material can be trapped in undesirable locations, necessitating surgical removal.

Candidacy

Who is a candidate for a hearing aid? The simple answer to this question is anybody with a confirmed hearing loss who is experiencing difficulty hearing. Unfortunately, the answer is not quite so straightforward. The best criteria for candidacy remain lifestyle, level of frustration, audiometric confirmation of a hearing loss, and a high level of motivation.

The next question relates to prescribing and fitting the most appropriate hearing aid. Because there are numerous models, a decision has to be made as to which type will provide the highest level of benefit to the patient. Although the major determinant should always be the extent and nature of the hearing loss, the cosmetic concerns of the individual also must be taken into account. Digitally programmable and fully digital hearing aids offer better opportunities to a greater segment of the population, but they are not without exception the only rational choice. An aid with more usual circuitry may be better able to serve the auditory and financial needs of the patient because digital technology can be up to 2.5 times more expensive than its analog counterpart.

A related issue is binaural fitting: in the majority of cases, two (bilateral) hearing aids provide better fidelity, speech understanding in noise, and localization ability than does one (unilateral) hearing aid. Exceptions to this rule are individuals with unilateral and/or markedly asymmetric hearing losses. The benefits of two hearing aids should be clearly and thoroughly explained to patients, although some will opt to obtain one aid at first and a second aid to follow. Because many insurance companies do not pay or reimburse for hearing aids, finances as well as psychological barriers may play a significant role in the decision-making process. The dispensing audiologist needs to be cognizant and respectful of the patient's economic situation and do whatever is possible to alleviate the strain of the circumstance. In summary, in addition to the nature and extent of the hearing loss, social and economic factors constitute an equal part of the equation in the prescription of hearing aids.

As with any prosthetic device, patient and family expectations play a significant role in the ability to adapt to, and make maximum use of, a hearing aid. During the initial evaluation phase, the patient should be advised that sound via a hearing aid is not the same as normal hearing. Nor will the adjustment to the new sound necessarily be either rapid or easy. The sound quality does not mirror that of the normal auditory system, and when speech discrimination is poor, a hearing aid may not address the problem well enough to satisfy the needs of the individual. Furthermore, the lack of ability to understand speech in noisy situations is one of the most frequent complaints of the new hearing aid user. In fact, the ability to overcome background noise and deliver an intelligible speech signal remains one of the most challenging technological problems facing hearing aid researchers.

Hearing aid users often require additional assistance in certain listening situations, usually when background noise prevents recognition of the principal input stimulus. Additionally, there are those with mild hearing losses who function well under most listening conditions but experience difficulty in specific instances, including movies, lectures, and conferences. Assistive listening devices can facilitate listening under these difficult circumstances. These aids provide sound, television, and telecommunications enhancement and signal-alerting capabilities. Various types of systems exist: frequency modulation (FM), induction loop, infrared, and a hard-wired amplification device. The decision as to the most appropriate type of device depends on the specific needs of the individual, geographical requirements, and environmental surroundings. The devices can be coupled to the hearing aid, or, for those who do not use an aid, they are stand-alone amplification systems, which provide benefit in a specific listening situation such as television viewing. In addition, visual enhancement and alerting systems including closed captioning and lights can be used as a substitute for alarms, doorbells, telephones, and so on. Informing and counseling the hearing-impaired individual on the types and benefits of these very useful devices can serve to improve both the quality and safety of daily living.

BONE-ANCHORED HEARING AIDS

The use of bone-anchored hearing aids (BAHAs) started 20 years ago in Sweden, and there are now close to 7500 patients worldwide who have been implanted and fitted successfully with this device.

The general idea of a convention bone conduction hearing device is that the bone-conducted sound bypasses the impaired and/or diseased external or middle ears. With BAHA, there is direct bone conduction, without skin and soft tissues being part of the vibration transmission path. The system is composed of the fixture and a bone-anchored abutment (**Fig. 31–1A**) that is placed during surgery, and the sound processor that can be either an ear-level device (**Fig. 31–1B**) or a body-worn aid that is fitted 3 to 4 months after surgery.

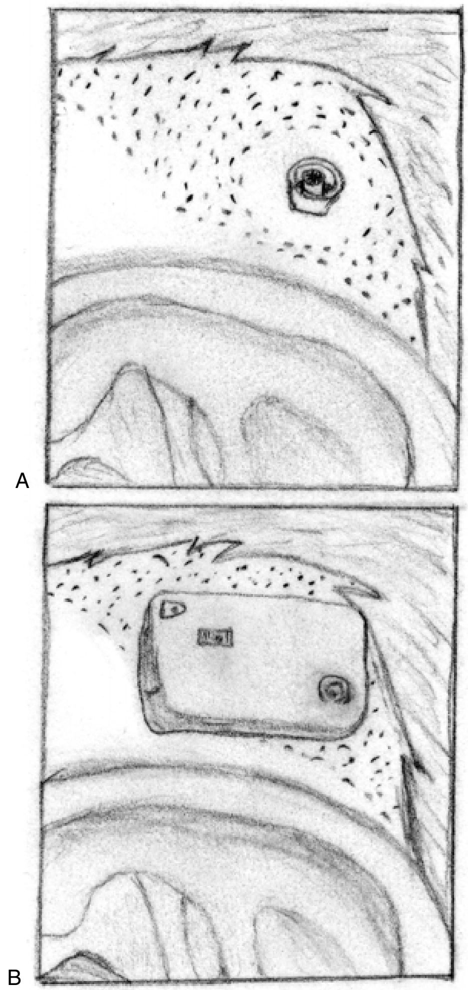


Figure 31–1 ***(A)*** *Bone-anchored hearing aid (BAHA) abutment fixed in place and ready to be connected to a processor unit.* ***(B)*** *An ear-level BAHA processor unit attached to the abutment.*

Candidacy

Overall, the cochlear hearing threshold (reserve) should be better than 45 dB for the BAHA ear-level device (**Fig. 31–1B**) and not worse than 65 dB for the body-worn aid. The size of the air–bone gap is of no significance because the BAHA bypasses the ossicular chain. BAHA candidates are patients who have a conductive or mixed hearing loss and who can still benefit from sound amplification.

Some particular indications for BAHA candidacy are the following:

- Single-sided deafness, where BAHAs can offer improved speech recognition in noise and reduction of the head shadow effect
- A chronically draining ear, where the use of an air conduction hearing aid aggravates the infection, causes a feedback problem, causes poor wearing

comfort, or poor sound quality (In fact, patients with recurrent otitis externa, draining otitis media, or a radical mastoidectomy cavity can all benefit from a BAHA.)

- Congenital ear canal malformations where the cochlea is intact and functional

There are a few contraindications for the use of a BAHA to keep in mind:

- Poor hygiene (in children, the responsibility falls to the parents)
- For patients in the United States, age of less than 5 years
- If sufficient bone volume and bone quality is not present for the successful anchoring of the BAHA abutment within a patient's skull

Medical and Audiological Evaluation

Medical evaluation to check a patient's hygiene and for presence of any disease affecting the skin of the scalp is required when assessing a patient for a BAHA device.

Audiolological preoperative measurements are acquired by pure-tone audiometry and speech audiometry (a score of better than 60% phonetically balance (PB) word list is recommended).

The test rod is performed with a plastic bar to which a bayonet or snap coupling is attached at one end. It is intended to assess the candidate preoperatively, to educate the patient, and to demonstrate to the prospective patient the expected result.

Surgical Procedure

A skin incision is performed using the BAHA dermatome (or manually); a skin flap is made with the subcutaneous tissue cut down to the periosteum. The periosteum at the fixture site should be incised and removed from the bone, and the fixture site is prepared using the surgical guide for drilling. This drilling guide indicator is also used during the actual tapping and fixture insertion. A cover screw placement may be used to protect the inner hole of the fixture temporarily if a two-stage procedure is planned. In this case, the fixture is left for a period of 3 to 6 months, during which time osseointegration takes place before the abutment can be fitted to the BAHA unit.

There are two prerequisites for establishing and maintaining a reaction-free skin penetration. The skin surrounding the fixture should be hairless to help keep the fixture site clean, and the skin flap must be very thin to avoid any movement of skin around the abutment. After finalizing the soft tissue preparation, a hole is punched over the fixture site with a 4 mm biopsy punch. The abutment is then correctly fitted to the fixture. Finally, the healing cap is attached to the abutment to fix the dressing in place and prevent formation of a hematoma. The fixture has to osseointegrate with the bone for 3 months before fitting the BAHA sound processor. A standard mastoid dressing is left in place for 1 or 2 days. Seven days after the healing cap is removed, a new dressing is placed for 7 days. Patients with two-stage surgery can be fitted 1 month after the second stage. Patients with one-stage surgery are fitted after 3 months. *Warning*: Early loading may result in loss of the fixture.

Follow-Up

A daily cleaning routine is very important to maintain the integrity of the site and to prevent a reaction with the skin. A follow-up program of twice-a-year inspection of the site is sufficient; the skin and the stability of the abutment can be checked at those times.

The fixture and the abutment can be left in place if the patient has to undergo a magnetic resonance imaging (MRI). scan However, the sound processor unit should be removed prior to the MRI procedure.

COCHLEAR IMPLANTS

Despite the extent of the development of new hearing aids, a critical problem remains for those children and adults with severe to profound sensorineural hearing loss who receive little or no benefit from amplification by either conventional hearing aids or BAHAs. The purpose of the cochlear implant is to bypass the traditional form of sound transmission and transduction by using direct electrical stimulation of the auditory neurons and the auditory nerve. No matter how sophisticated a hearing aid, its basic mode of operation is still the amplification of incoming sound. In contrast, a cochlear implant attempts to replace the function of the auditory hair cells that has been lost by the damaged cochlea. In a normal-hearing ear, the hair cells within the cochlea act as a transducer of mechanical energy into electrical energy capable of stimulating a patterned discharge from the eighth cranial nerve (CN VIII). An actual decrease in the number of hair cells or in hair cell function, resulting in a significant hearing loss, causes the cochlea to lose its ability to execute the transduction function that results in CN VIII nerve stimulation and consequently hearing (see Chapter 26). The implant replaces the function of the lost hair cells by converting

Figure 31–2 *Three external devices from one of the three cochlear implant manufacturers (i.e., Tempo+) approved by the U.S. Food and Drug Administration (FDA), showing a selection of two behind-the-ear processors and one processor to be worn on the body (e.g., on the belt).*

the mechanical energy of sound into electrical energy capable of exciting the CN VIII. Cochlear implants are placed within the cochlea to stimulate directly auditory neurons in Rosenthal's canal and the CN VIII nerve endings.

Technology

Historical Aspects

In 1957, Djourno and Eyries stimulated a cochlear nerve exposed during surgery. An electrode was placed directly on the nerve, and the exposed nerve was stimulated with a simple electric current that produced the perception of an auditory sensation in the patient. This experiment led to the concept of direct stimulation of the auditory nerve, which is the basis for cochlear implant function. Following this report, additional investigations led to the development of a speech processor to interface with and drive an electrode implanted in the cochlea. In 1972, the House 3M was the first commercially marketed single-channel cochlear implant. The first multichannel device was developed in 1982 at the University of Melbourne, Australia. Since then, several other manufacturers have produced and marketed versions of multichannel cochlear implants.

Structure of a Cochlear Implant

Devices presently approved by the U.S. Food and Drug Administration and being used in the Unites States consist of those manufactured by Advanced Bionics Corp. (Valencia, CA, United States), Cochlear Corp. (Melbourne, Australia), and MED-EL, Medical Electronics Corp. (Innsbruck, Austria). Although there are several types of cochlear implants, the fundamental construction consists of both an external component (**Fig. 31–2**) and a surgically implanted internal component (**Fig. 31–3**). The external portion of the device includes a microphone, a microprocessor-based speech processor (SP), and a radio frequency (RF) transmitting coil (**Fig. 31–2**). The implanted portion of a cochlear implant houses an RF receiver coil, microprocessor-based stimulator, and multichannel electrode array (**Fig. 31–3**).

The microphone detects sound and sends that information to the SP. The SP electronically processes the incoming sound based on programs stored in memory. Programs are

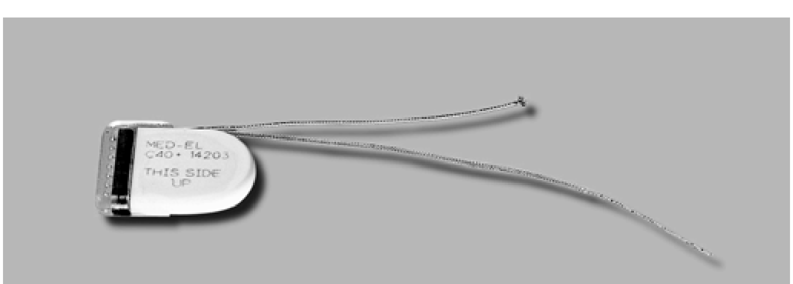


Figure 31–3 *The internal device from one of the three FDA-approved cochlear implant manufacturers (i.e., Combi40+).*

software algorithms that encode the information provided by the microphone and provide the instructions that enable the receiver/stimulator and electrode array to respond appropriately and to electrically stimulate a specified site within the cochlea. Programming of the cochlear implant will vary by device, and even for the same type of processor, depending on the specific CN VIII anatomy and the location of the remaining spiral ganglion neurons for each individual. Each cochlear implant device and model within a group of devices from the same manufacturer can use different algorithms to encode auditory information, and this has to be applied to individual patients, who will have their own unique requirements for electrical stimulation of the cochlear nerve and spiral ganglion neurons.

Once the signal has been processed and encoded by the SP, the information is sent to the transmitting coil, which is held in place above the implanted receiver/stimulator by an external magnet. The function of the transmitting coil is to transmit the signal provided by the SP to the receiver/stimulator via an RF signal. Upon reaching the receiver/stimulator, the signal is decoded and transduced into a series of electrical pulses for the different channels of the electrode array. Information encoded into the signal tells the receiver/stimulator which electrodes to activate, and when and how to activate the selected electrodes. The electrical current sent through the implanted electrodes to the cochlea directly stimulates remaining auditory neurons and CN VIII fibers. The patterns of stimulation are conducted along the auditory nerve to the auditory nuclei of the central nervous system and interpreted as meaningful sound.

Electrode Arrays

In the earlier years of development, implants had either a single or multiple electrodes and channels, and the surgical placement of the electrode array was either extra- or intracochlear. Currently, the designs are all intracochlear and have both multiple electrodes and multiple channels. The specific pattern of electrode stimulation depends on the approach used to convert the incoming acoustic stimuli into electrical impulses.

Recent research into electrode design has focused on the development of cochlear implant electrode arrays that will lie closer to the modiolus of the cochlea. Because of their closeness to the anatomical position of the spiral ganglion and the CN VIII, these electrodes will be able to run complex programming strategies with less power, opening the way for ear level and ultimately fully implantable SP devices.

Special electrodes have also been designed for ossified cochlea (i.e., compressed electrodes and split electrodes), and short electrode arrays have been designed to achieve electroacoustic bimodal stimulation (EAS) of both the auditory nerve and the residual hair cells in the apical portion of the cochlea. There is also an increase in the number of patients with residual hearing who are having cochlear implant surgery. Therefore, minimizing trauma to the inner ear by the electrode array is one of the focuses of cochlear implant research and electrode design. Histological evaluation of the cochlea after the insertion of a cochlear implant electrode in human cadaver temporal bones has demonstrated immediate damage to the spiral ligament, organ of Corti, osseous spiral lamina, and other fine structures of the inner ear. A histological study of human temporal bones analyzed the electrode position and structural damage to the cochlea and suggested a scale of 1 to 4 to describe a gradation of the extent and severity of the initial macroscopic trauma (**Table 31–1**). Electrode insertion experiments with fresh cadaver temporal bones have demonstrated that direct trauma to cochlear structures (**Fig. 31–4**) can be partially prevented by the design of less traumatic electrodes and modification of the surgical technique used for electrode insertion (e.g., decreasing the size of the cochleostomy used for insertion of the electrode array).

Recent animal studies (rats and guinea pigs) have characterized the electrophysiological pattern of hearing loss after cochlear electrode insertion trauma. Laboratory rats and guinea pigs were evaluated for hearing acuity before and after cochlear electrode insertion using distortion product otoacoustic emissions (DPOAEs) and auditory brainstem responses (ABRs). There were progressive increases in ABR thresholds and decreases in ABR amplitudes in the electrode insertion trauma animals. The amplitude of the DPOAEs in these experimental traumatized cochleas also showed progressive

Table 31–1 Grading system describing cochlear trauma caused during cochlear implantation

Grade 1	No observable macroscopic trauma*
Grade 2	Elevation of basilar membrane
Grade 3	Dislocation to scala vestibuli
Grade 4	Fracture of osseous spiral lamina or modiolus or tear of stria vascularis/spiral ligament

*Possible molecular damage leading to apoptosis. (Derived from Eshraghi AA, Yang N, Balkany T. Comparative study of cochlear damage with three perimodiolar electrode designs. Laryngoscope 2003;113(3):415–419.)

Figure 31–4 *A midmodiolar section of a fresh, unfixed temporal bone implanted with a cochlear implant electrode array showing a fracture of the osseous spiral lamina (large arrow) and displacement of the basilar membrane at 180 degrees (small arrow). The presence of the electrode in the scala vestibuli at 360 degrees indicates passage of this electrode through both the basilar membrane and the scala media. This temporal bone demonstrates grade 4 trauma (see Table 31–1). (Image modified from figure 3 of Eshraghi AA, Yang N, Balkany T. Comparative study of cochlear damage with three perimodiolar electrode designs. Laryngoscope 2003;113(3):415–419.)*

decreases. These results document a progressive loss of hearing acuity postimplantation for all frequencies and strongly suggest that electrode insertion trauma generates oxidative stress within injured cochlear tissues. Therefore, trauma-induced hearing loss caused by insertion of a cochlear implant electrode is thought to be caused by the combination of direct acute trauma to the inner ear structures (necrosis) and the delayed effect of trauma-initiated oxidative stress on the hair cells that survived the initial physical trauma (i.e., apoptosis of these damaged hair cells). The therapeutic implication of these findings is major because we can now expect that future electrode designs will be more atraumatic and that future implant design will incorporate a drug delivery system that can deliver otoprotective molecules to prevent progressive loss of hearing after implantation and/or regenerative drugs to stimulate repair of the cochlea and the CN VIII.

Speech Processor and Coding Strategies

The principal variables of processing strategies include the number of available channels, rate of stimulation, analog or pulsatile stimulation, simultaneous or sequential stimulation, speech feature extraction or waveform representation of incoming stimuli, and monopolar or bipolar stimulation. Currently, variations of several speech-encoding strategies are being implemented in commercially available cochlear implants, including continuous interleaved sampling (CIS), spectral peak (SPEAK), advanced combined encoder (ACE), simultaneous analog system (SAS), and high resolution (HiRes).

CIS employs a nonsimultaneous interleaved pulsatile form of stimulation where the pulses, although interleaved, are sequential and not overlapping to minimize electrode interaction, and can be delivered at high rates of stimulation. In SPEAK, six to 10 of the available electrodes are selected for stimulation based on the spectral peaks with the greatest energy of the incoming auditory stimuli. ACE assigns electrodes to be stimulated based on the spectral information (SPEAK) but incorporates more rapid and adjustable rates of stimulation to convey the timing information of the speech signal (CIS). SAS uses analog transmission, where the electrodes are stimulated simultaneously in a bipolar mode, and the HiRes strategy is capable of high rates of stimulation. In addition to a diversity of encoding strategies, current devices are equipped with highly flexible programming options, bidirectional communication links (telemetry), ear-level processors, and other device-enhancing features. Concurrently, electrode arrays are being designed to be less traumatic during insertion into the cochlea and to be more efficient signal transmitters.

Candidacy

Candidacy criteria for cochlear implantation have changed markedly since cochlear implants first emerged as a viable treatment for deafness. During the early years, data were obtained from postlingual profoundly deaf adults, and following that, postlingual profoundly deaf children older than 2 years. These restrictive candidacy criteria were based on a general lack of knowledge and experience as well as a scarcity of data to confirm that a substantial benefit was derived from cochlear implants. As the number of recipients increased and devices became more sophisticated and flexible, speech understanding with the various cochlear implants rose rapidly and has far exceeded the initial expectations.

The reported increased ability to comprehend speech by implant recipients, due mainly to improvements in technology, has provided the impetus to expand the population eligible for cochlear implantation. Current candidacy criteria include the geriatric population, adults and children with congenital (prelingual) long-term deafness, children ages 12 months and older, and individuals with multiple disabilities. In fact, because newborn hearing screening programs have led to earlier identification of hearing loss, implantation of hearing impaired children under 12 months of age is now being done under certain circumstances.

During the 1980s candidacy criteria included "no benefit from amplification," which was defined as a score of 0% on monosyllabic words under the best-aided condition. As the level of speech understanding with cochlear implants increased to an average of ~90% on sentence identification in postlingual deaf adults, the candidacy criteria expanded to those adults and children who possess increasing amounts of residual hearing. Results of studies have indicated that children and adults with residual hearing performed significantly better with cochlear implants than with hearing aids. Currently, prospective pediatric and adult recipients who have substantial sentence recognition abilities with amplification are receiving cochlear implants and are achieving substantial gains postimplantation.

Other areas that are gaining increasingly widespread research and clinical attention include bilateral implantation of adults and children and electrical/acoustic bimodal stimulation of patients with residual hearing in the low-frequency portion of their cochleas.

Prior to implantation, children and adults undergo extensive age-appropriate audiological, otologic, radiological, and communication skills evaluations both for candidacy determination and to establish baseline evaluations by which to measure the outcomes gains achieved by cochlear implantation. In addition, age-appropriate phoneme, word, and sentence recognition tests are administered preoperatively and postoperatively at regular intervals to monitor implant performance and gains for the patient.

Surgical Procedure

Cochlear implant surgery is performed under general anesthesia, with perioperative antibiotics. A postauricular incision, with a sometime posterior/superior extension, is made, and a scalp flap is elevated. The pericranium is elevated, and the previously marked well to house the electronics package is drilled, usually down to the dura in young children. A complete mastoidectomy is then performed, and a channel is drilled between the well created for the electronics and the mastoid cavity. The posterior mesotympanum is approached through the facial recess, and a cochleostomy is drilled just anterior to the round window membrane niche. The electrode is then inserted ~25 mm into the scala tympani, the device is fixed to bone with nonabsorbable sutures, the proximal lead is fixed in some manner to bone, and the wound is closed after device function has been checked electronically. An x-ray can be taken to document electrode position, and then a mastoid dressing is applied. The patient typically is discharged the morning following the surgery.

The five most serious complications associated with cochlear implantation involve the facial nerve, electrode placement, flap necrosis, infection, and migration or extrusion. It is difficult to determine exact complication rates because the manufacturers can only calculate what is reported to them. As of July 1998, Cochlear Corp. reported a complication rate of 3.5% on a total of 283 children and adults implanted with the CI 24 M, which was undergoing clinical trials. Advanced Bionics reported a 4% complication rate for a total of 783 children and adults who participated in their clinical trials. The Nucleus 22, the precursor to the CI 24 M, had a 1.4% complication rate on a population of 5170 adults and a 0.85% complication rate on a group of 4051 children. Recently, complication rates have decreased in the population implanted because surgeons have more experience and the designs of the implants have improved along with the manufacture of the implants (i.e., lower device failure rates).

Outcome Assessment

During the 1980s and early 1990s, postoperative expectations for children were limited to the auditory perception of suprasegmental features of the speech signal and closed-set speech recognition. The impact of these and other influencing factors has been the development of significant amounts of open-set speech perception in pediatric cochlear implant recipients. Following several years of implant usage, average open-set monosyllabic word scores reported by several investigators in the early to mid-1990s ranged from 0 to 33%; more recent published results reveal mean word score levels of 75% and higher. It is important to realize, and account for, however, the wide range of results reported in the literature. Notwithstanding the substantial improvements in postimplantation performance and the increasing numbers of adults and children who achieve better scores, there still remains a great amount of variability in outcome across all devices that are implanted. As we expand the criteria for implantation, we should not lose sight of the many variables that still affect results and, although they may be modified over time, will continue to be factors in the foreseeable future. The variables listed following here are not inclusive but have been shown to be some of the most important determinants.

In particular, two factors have been shown to play a significant predictive role in postoperative speech understanding: advancements in technology, including speech processing strategies and electrode design, and the length of profound deafness in the patient prior to implantation. Other confounding elements are survival and overall health of the auditory neurons and CN VIII elements, age at time of implantation, medical/surgical issues, device programming expertise, length of implant usage, counseling and expectations, and, particularly in children, communication mode, education, and type and frequency of intervention.

Despite the many questions that remain regarding the factors that can influence auditory perception performance, little doubt exists as to the efficacy and safety of pediatric cochlear implantation. As a natural consequence of the excellent perception results, current studies are assessing the extent of development of oral language and speech production skills following implantation. Recent studies have shown that implanted children acquire language at a rate equal to that of normal-hearing children (12-month growth in 12 months) in contrast to nonimplanted deaf children who acquire linguistic skills at half the rate of normal-hearing children (6-month growth in 12 months), and some studies have even reported implanted children developing language at a rate that exceeded that of normal-hearing children (9-month growth in 6 months). Although language skills improve in all implanted children, those children who are trained to use oral communication perform at a higher level than those children who use total communication. The same relationship has been found for speech production skills: implanted children with good perception skills using oral communication have more intelligible speech than those children who use total communication and/or have poorer perceptual skills with the implant.

Postimplantation speech perception results in the adult population have undergone a similar evolution. Prior to the current processing strategies, monosyllabic word scores reportedly averaged 22%, and sentence scores averaged 40%. More recent mean word and sentence recognition scores with the newer processing strategies have reached increasingly higher scores, with a significant portion of the adult population achieving scores of 80 to 100% postimplantation. Despite these very impressive results that have allowed implanted adults to communicate in personal and professional situations and use the telephone, there still exists a wide range of performance that underscores the need to continue to explore variables that influence outcomes in adults.

SUMMARY

The design and stimulation strategies of cochlear implants and the procedures used during cochlear implant electrode insertion have come a long way over the past 40 years. They have evolved from single-electrode, single-channel devices to sophisticated multichannel, multi-electrode devices offering multistrategy stimulation flexibility. The field continues to be dynamic, with concurrent research focused on totally implantable devices, electrode design to facilitate transmission, and improved processing schemes and other innovations, such as the bimodal stimulation approach of electro-acoustic stimulation (EAS). EAS uses a dual approach employing both cochlear implant stimulation of the damaged high-tone area of the cochlea and hearing aid amplification for the stimulation of any residual low-tone hearing present within the apical portion of the patient's cochlea. Advances in the research of the molecular biology of cell death pathways suggest that incorporating oto- and neuroprotective agents (e.g., inhibitors of an involved cell death signal cascade) during and after the procedure of electrode insertion will help to conserve the remaining sensorineural elements (i.e., both hair cells and neurons) within the cochlea, leading to conservation of residual hearing, and will therefore result in an enhancement of cochlear implant performance and retention of the patient's residual hearing capacity. Future advancements in technology and otoprotection strategies to conserve the remaining sensorineural elements of the cochlea have the possibility of providing superior performance to current implant users and to address the needs of a wider segment of the hearing-impaired population.

Suggested Readings

Balkany T, Hodges A, Eshraghi AA, Butts S, King J. Cochlear implants in children: a review. Acta Otolaryngolog (Stockh) 2002;122:356–362

Dillon H. Hearing Aids. New York: Thieme Medical Publishers; 2001

Djourno A, Eyries C. Prothèse auditive par excitation électrique a distance du nerf sensoriel a l'aide d'un bobinage inclus à demeure. Presse Med 1957;35:14–17

Eshraghi AA, Hodges A, Telischi F, Balkany T. Cochlear implant technology. In: R. Jackler and D. Brachman, eds. Neurotology. Philadelphia: Elsevier; 2005,1301–131A

Eshraghi AA, Yang N, Balkany T. Comparative study of cochlear damage with three perimodiolar electrode designs. Laryngoscope 2003;113(3):415–419

Scarpidis BS, Madnani D, Fletcher C, et al. Arrest of apoptosis in auditory neurons: implications for sensorineural preservation in cochlear implantation. Otol Neurotol 2003;24(3):409–417

Waltzman S, Cohen N. Cochlear Implants. New York: Thieme Medical Publishers; 2000

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. The best criteria for hearing aid candidacy include
   A. A confirmed hearing loss
   B. Lifestyle
   C. Patient motivation
   D. All the above
   E. A and C
2. Binaural hearing aids benefit
   A. All those with a hearing loss
   B. Only those patients with sound localization problems
   C. All patients except those with unilateral and asymmetric hearing losses
   D. All the above
   E. B and C
3. The principal variables of speech-encoding strategies include
   A. Analog versus pulsatile stimulation
   B. Monopolar versus bipolar stimulation
   C. Simultaneous versus sequential stimulation
   D. All the above
   E. A and C
4. Average open-set sentence recognition scores in adults using the most current encoding strategies exceed
   A. 30%
   B. 40%
   C. 50%
   D. 60%
   E. 80%

Chapter 32

MECHANISM OF NOISE-INDUCED HEARING LOSS AND OTOPROTECTIVE STRATEGIES

RICHARD D. KOPKE, JOHN K.M. COLEMAN, JIANZHONG LIU, RONALD L. JACKSON, AND THOMAS R. VAN DE WATER

Despite continuing advances in hearing conservation over the past several decades, noise-induced hearing loss (NIHL) remains a major cause of deafness in industrialized nations. Significant reasons for the persistence of this problem include steadily increasing noise levels and inherent limitations to environmental engineering solutions and personal hearing protection devices (HPDs). Recently, insights into the molecular mechanisms of noise-induced cochlear injury have led to new treatment strategies that render the cochlea more resistant to noise as well as enhance the recovery of noise-injured cochleae. Combining this therapeutic approach with current hearing conservation practices will improve treatment of and protection from acute NIHL. Future work may focus on cochlear sensory cell regeneration to treat and restore hearing even after chronic hearing losses.

The U.S. military and the National Institutes of Health (NIH) have done much to promote and standardize hearing conservation practices over the last several decades. Despite advances, NIHL is still one of the most common military disabilities, affecting an estimated 10 to 15% of the armed forces. Significant hearing threshold shift (STS) rates vary from 29% to over 70% in some at-sea Navy job specialties. STS rates tend to be even higher for aircrew, whether land or sea based. In several studies, STS rates were reported to be as high as 11% after only short periods of periodic weapons-generated noise exposure.

Besides the military, many other vocations are considered to be noise hazardous. The National Institute of Deafness and Other Communication Disorders (NIDCD) estimates that some 10 million persons in the United States have some degree of NIHL. The railroad, plumbing, carpentry, and coal-mining industries are but a few of the noise-hazardous occupations. Recreational firearm use is popular in the United States and contributes significantly to the overall problem, and NIHL contributes significantly to the 30 million individuals in the United States with hearing impairment and to the associated \$56 billion annual cost. In addition, NIHL is considered to be an international problem, with some 600 million individuals worldwide estimated to be at risk primarily through occupational exposure.

Environmental engineering solutions to abate noise are a critical part of prevention and are often effective; however, they generally are impractical, economically unfeasible, or ineffective due to physical and acoustic parameters. Another important line of defense is the use of HPDs during periods of sound exposure. Lack of compliance obviously negates their effectiveness for the protection of hearing; moreover, HPDs suffer from several inherent limitations. Often, the level of noise exceeds the protective capability of the HPD, or damaging sound energy is transmitted directly through the skull, bypassing the protective device. Maintaining a comfortable and reliable acoustic fit of an HPD can be problematic. The element of surprise coupled with the need to periodically remove the device to hear also can lead to acoustic injury. NIHL can occur following even a short exposure to intense sound. Furthermore, evidence suggests that inhaled toxicants commonly found in noisy occupational environments, such as carbon monoxide, ethyl benzene, toluene, and styrene, can act additively to injure the noise-exposed inner ear. Many of the environmental- and HPD-limiting factors are difficult to overcome; hence, pharmacological-based therapies to render the cochlea more resistant to noise damage, as well as reverse or treat acute noise-induced hearing loss, may prove useful.

HOW ACOUSTIC OVEREXPOSURE DAMAGES THE COCHLEA

To develop a rational approach to pharmacological therapies for the prevention or treatment of acoustic overexposure, an understanding of the mechanisms by which noise injures the cochlea must be understood. Over the past decades, significant progress has been made in understanding the cellular and molecular mechanisms of noise injury to the cochlea. This, in turn, has led to exciting experimental data, which appear promising for the development in the near future of both protective and therapeutic agents to ameliorate the effects on sound trauma.

Cochlear injury due to excessive noise can be divided broadly into mechanical and metabolic mechanisms. With exposures of ~115 to 125 dB sound pressure level (SPL) at the ear, mechanical damage tends to predominate. This damage may include the disruption of such structures as Reissner's membrane, basilar membrane–cell junctions, damage to or loss of stereocilia bundles, and even disruption of subcellular organelles, such as the endoplasmic reticulum. Damage to hair cell–Deiters' cell junctions at the level of the reticular lamina can lead to an admixture of potassium-rich endolymph, with the perilymph surrounding the cell bodies of the outer hair cells, leading to their destruction. However, in most clinically relevant scenarios, the level of noise exposure to the ear is less than 115 dB, and the damage tends to be metabolically driven. (The remainder of this chapter will focus on metabolic causes of noise-induced injury of auditory sensory cells and how they may be ameliorated.)

The cochlea is a highly metabolically active sensory organ, which receives 0.5 mL per minute of blood flow under normal conditions, a relatively high amount of flow compared with other organs of similar size. One theory of noise-induced cochlear injury is that a consequence of acoustic overexposure is a temporary reduction in blood flow leading to a cochlear ischemia-reperfusion injury similar to that which occurs during a stroke in the brain.

Although several earlier studies have suggested that cochlear blood flow (CoBF) may increase with some excessive noise exposures, the evidence from many recent studies is that CoBF actually decreases to the cochlea during loud sound exposure. Investigations utilizing intravital microscopy, which allows for real-time, continuous and quantifiable observation of the vessels of the lateral wall of the cochlea, have demonstrated localized ischemia in response to prolonged exposure to a loud sound. In addition, researchers using laser Doppler flow meters have made measurements of blood flow in the cochlea during loud sound exposure that are consistent with the findings of the other reports, that CoBF decreases during exposure to a traumatic level of noise.

The current development of this theory is that noise provokes hypoperfusion and ischemia in the microcirculation of the cochlea, followed by reperfusion of the cochlea, and this generates reactive oxygen species (ROS) and other free radicals, as has been demonstrated, to occur in the brain in response to a stroke. ROS have been shown to form in the cochlea after either loud sound exposure or blast trauma. It is believed that these ROS can assault the sensory cells of the cochlea, which can lead to cell death, and thus the permanent threshold shifts associated with NIHL.

Free radical molecules are characterized by the presence of unpaired electrons, making them chemically unstable and highly reactive with cellular proteins, membrane lipids, deoxyribonucleic acid (DNA), and cellular organelles. ROS are further characterized in that they contain at least one molecule of oxygen. Evidence for the production of superoxide and hydroxyl radical anions in the cochlea in the face of noise-induced ischemia-reperfusion has been published.

OXIDATIVE STRESS HYPOTHESIS FOR NOISE-INDUCED COCHLEAR INJURY

Evidence has subsequently accumulated that indicates that ROS play a major role in certain acoustic overexposure conditions. The evidence for the oxidative stress hypothesis is strongly supported by published work from several different research centers. **Fig. 32–1** depicts a summary of oxidative stress in the cochlea, depicting free radicals that may be generated (**Fig. 32–1A**), intrinsic cochlear antioxidant defenses (**Fig. 32–1B**), and the consequences of oxidative damage to hair cells when the antioxidant defenses of the auditory sensory cells are overwhelmed (**Fig. 32–1C**). Some of the intrinsic cochlear antioxidant defenses include antioxidant enzymes, heat shock proteins, trophic factors, small molecules such as vitamins C and E, and the crucial antioxidant molecule, reduced glutathione (GSH), as shown in **Fig. 32–1B**.

Acoustic overexposure leads to a modulation of antioxidant enzymes and a reduction in GSH levels within the cochlea. The generation of ROS is correlated with acoustic overstimulation in the cochlea, even after the cessation of the noise exposure. If these ROS are infused without any noise exposure into the cochlea, they produce a characteristic pattern of damage and hair cell loss just as if the cochlea has been overstimulated by a damaging level of sound. Lipid peroxidation, one of the hallmarks of oxidative stress injury within a cell, has been observed in the spiral ganglion, organ of Corti, and stria vascularis, after acoustic overstimulation. A reduction of GSH availability to the inner ear intensifies noise-induced cochlear injury. Conversely, augmenting inner ear antioxidant defenses by increasing antioxidant enzyme activity, increasing the level of available inner ear GSH, or adding exogenous antioxidant compounds can reduce

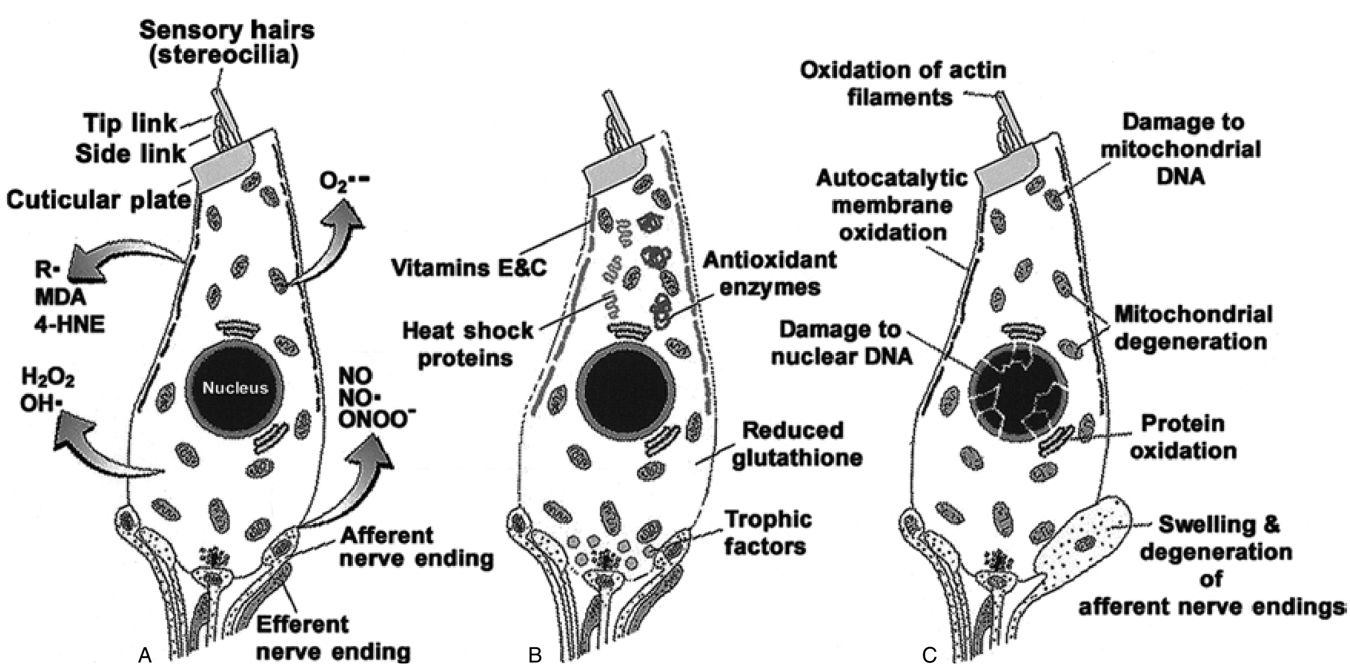


Figure 32–1 *Noise exposure causes oxidative stress injury to cochlear sensory cells.* ***(A)*** *The main types of free radicals and reactive oxygen species (ROS) demonstrated or postulated to be formed in the cochlea during acoustic overexposure.* ***(B)*** *The hair cell's defenses against oxidative stress.* ***(C)*** *When free radicals and ROS overwhelm the oxidative defenses of a cell, damage occurs to numerous cellular components, leading to cell injury and/or cell death. 4-HNE, 4-hydroxy-2,3-nonenal; MDA, malondialdehyde.*

the amount of permanent hearing loss caused by noise. It has been reported that the well-known increased sensitivity to noise and other toxins of the basal region of the cochlea may in part be due to a relative weakness of antioxidant defenses [i.e., GSH in outer hair cells (OHCs) from basal regions as compared with the apical region] in this segment of the cochlear duct.

When ROS or free radicals are generated in excessive amounts, as may occur with acoustic overexposure, they overwhelm cochlear antioxidant defenses. These highly reactive compounds oxidize cell membrane lipids, intracellular proteins, and DNA, leading to injury and/or cell death, as shown in **Fig. 32–1C**. During acoustic overstimulation, ROS may be generated from a variety of sources, including the mitochondria or secondary to ischemia-reperfusion, toxic effects of release of excessive levels of glutamate, large increases in intracellular calcium, or microlesioning of the cell membrane. Excessive generation of ROS by the cell's mitochondria injures those mitochondria, which leads to further generation of ROS.

Molecules of ROS are unstable, and they interact with cell membranes, disrupting their integrity. Peroxidation of cell membranes initiates an autocatalytic chain reaction of cell membrane oxidation with the production and release of toxins such as 4-hydroxy 2,3-nonenal (HNE), an aldehyde adduct of membrane lipid peroxidation. HNE is itself a toxic molecule, and when applied to auditory sensory cells, it causes their death via apoptosis (programmed cell death). Protein oxidation may result in the loss of structurally important actin filaments, leading to the loss of functional hair bundles and the disruption of other important regulatory proteins that control maintenance of cell stability for critical ions (e.g., calcium). This can lead to further cell injury and culminate in the initiation of cell suicide; that is, apoptosis.

Lost hair cells are replaced by the expansion of the luminal surface area of neighboring supporting cells, leaving a scar as part of a healing process. However, these supporting cell scars that replace the auditory hair cells cannot transduce sound. It may take several days or even weeks after noise exposure until the actual loss of hair cells occurs. It is becoming apparent with narrowband noise exposures that a narrow segment of the organ of Corti correlated with the frequency of the injurious sound loses OHCs within a day or two. Then, depending on the intensity of the acoustic overexposure, waves of OHC losses occur in both basal and apical directions from the site of the initial injury over a period of days to weeks, much like the cell death that occurs in the areas of the brain (penumbra) that surround the initial site of a stroke lesion. The primary mode of OHC death in the apical- and basal-directed delayed losses appears to be apoptosis. This critical time interval can allow for a potential therapeutic window where rescue and repair could be initiated after the noise injury, but prior to the initiation of programmed cell death, thereby reducing the amount of permanent hearing loss.

ADDITIONAL GENERATORS OF OXIDATIVE STRESS DURING ACOUSTIC OVEREXPOSURE

If oxidative stress plays a major role in noise-induced cochlear injury, as current evidence suggests, then it is important to understand the potential generators of oxidative stress in more detail to formulate a rational approach to the development of effective pharmacological protection and treatment strategies. Although ischemia-reperfusion undoubtedly plays a role, evidence is accumulating that there may be other important sources of ROS and other free radicals in the noise-challenged cochlea. Postulated oxidative stress generators include mitochondrial generation of excessive ROS in cochlear cells, glutamate excitotoxicity secondary to excessive release of this neurotransmitter from afferent hair cell synapses during acoustic overexposure, GSH depletion in cochlear cells, and harmful fluxes in calcium ions within cochlear cells. It is likely that the resultant damage caused by cochlear oxidative stress is the result of multifactorial mechanisms acting in concert. This concept is summarized in **Fig. 32–2**.

MITOCHONDRIA

A major source of free radicals generated in oxidative-stressed cells is the mitochondria. Normally, 1% of the reactions of the electron transport chain yield free radicals. Under stress conditions, this percentage is increased, and when mitochondria are damaged, oxidative phosphorylation is inefficient, resulting in the production of even higher levels of ROS. One of the early subcellular changes noted in noise-stressed cochlear hair cells is mitochondrial swelling, a hallmark of mitochondrial oxidative injury. It can be seen as early as 2 to 4 hours after the onset of acute excessive noise exposure; hence there is evidence that (1) mitochondrial injury and (2) consequent enhanced levels of oxidative stress due to an uncoupling of oxidative phosphorylation are early events in the pathogenesis of acoustic trauma. In further proof, inhibition of mitochondrial self-repair intensifies

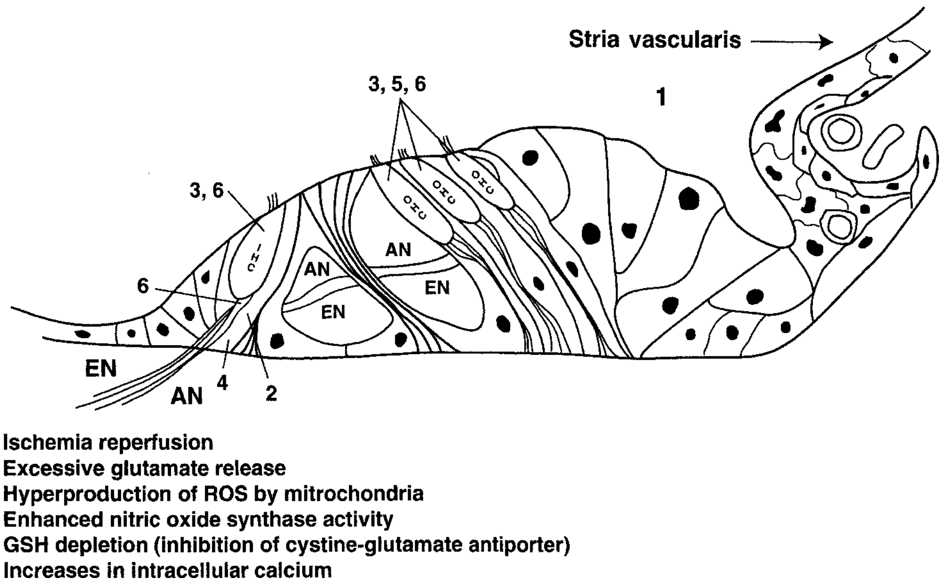


Figure 32–2 *Postulated generators of noise-induced oxidative stress within the cochlea. Diagram of the organ of Corti and stria vascularis depicting possible generators of oxidative stress occurring in response to acoustic overexposure. Possible sites of action: (1) ischemia-reperfusion injury may be generated in the stria vascularis by noise-induced vasoconstriction; (2) excessive glutamate release may occur at the inner hair cell afferent synapse and possibly at the small number of afferent synapses at the bases of outer hair cells (OHCs); (3) mitochondria production may lead to an abnormally high amount of reactive oxygen species (ROS) in hair cells; (4) increased nitric oxide synthase activity may lead to the production of high levels of nitric oxide and related free radicals; (5) depletion of glutathione (GSH) may occur predominantly within OHCs if the cystine-glutamate antiporter is inhibited; and (6) increases in intracellular calcium may occur in hair cells through a variety of mechanisms. AN, afferent nerve; EN, efferent nerve; IHC, inner hair cell.*

noise-induced inner ear injury. In addition to being damaged by the very free radicals that the injured mitochondria overproduce, other molecular events related to intense noise may lead to mitochondrial injury.

Glutamate excitotoxicity, GSH depletion, excessive increases in intracellular calcium, and ischemia-reperfusion can all lead to mitochondrial injury, and all of these factors have been implicated in noise injury to the cochlea. The consequences of mitochondrial injury include loss of key mitochondrial molecules, such as carnitine and cardiolipin, reduced activity of cytochrome c and other important enzymes, mitochondrial "electron leak" from the electron transport chain, loss of mitochondrial membrane integrity, and eventual onset of mitochondrial-induced cell death. Mitochondrial membrane permeability (MMP) is central in the cell death process. With a sufficient degree of mitochondrial injury, the mitochondrial membranes become permeable to molecules and release respiratory enzyme molecules (e.g., cytochrome c) that activate cell death effector proteins (e.g., caspases), which is one pathway for activation of apoptosis. Programmed cell death pathways involving calpain, caspases, and JNK/c-Jun molecules also have been noted to be activated in the cochlea in response to a damaging level of noise exposure.

ROS damage can depolarize mitochondria and cause an increase in the permeability of mitochondrial membranes that leads to pore formation and release of cytochrome c from the inner mitochondrial membrane. Cytochrome c, once released from its membrane-anchored location, can then pass through the pores formed in the outer mitochondrial membrane into the cytoplasm of the ROS-damaged cell. Once cytochrome c enters the cytoplasm, it enters into the apoptosome complex, where it combines with APAF-1 (the human homolog of the *c. elegans* proapoptosis molecule CED-4) and can now activate procaspase-9 into its active form, caspase-9 (**Fig. 32–3**). Activated caspase-9 can now act on downstream procaspase molecules, such as procaspase-3, to convert them to activated effector caspases (i.e., caspase-3). It is these activated effector caspase molecules that act on apoptotic substrates within the ROS-damaged cell that cause the degradation of membrane lipids, cellular proteins, and the typical pattern of DNA degradation that forms a characteristic pattern of DNA laddering that can be seen on gels and is considered to be a sure sign of a cell's irreversible

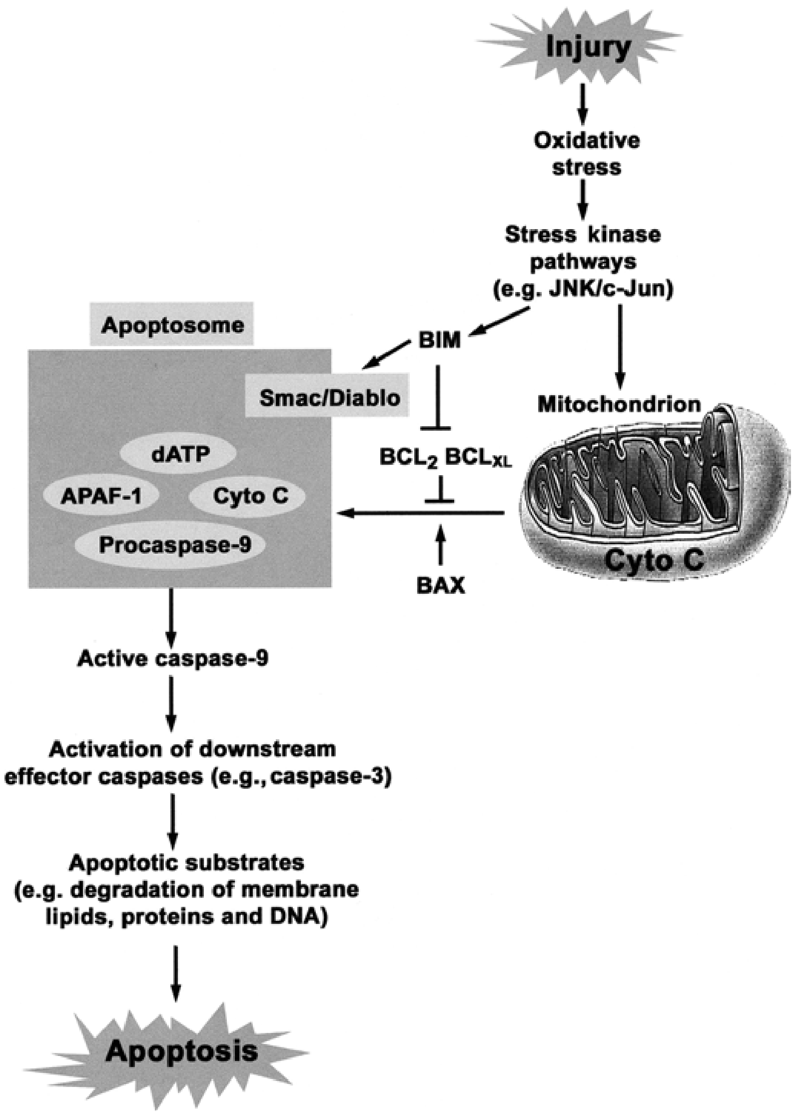


Figure 32–3 *Flow diagram of a cell-death pathway showing the cascade of events from initiating injury to apoptosis of damaged inner ear sensory cells. An injury generates oxidative stress within a sensory cell, causing the generation of ROS and free radical damage that activates a stress kinase cell-death pathway (e.g., mitogen activated protein kinase) [MAPK], that involves c-Jun-N-terminal kinase (JNK) and c-Jun. JNK/c-Jun can then cause both the activation of a proapotic BH-3, only member of the Bcl-2 family (BIM) and damage to both the inner and outer mitochondrial membranes that results in the release of cytochrome c into the affected cells cytoplasm. Once in the cytoplasm, cytochrome c enters the apoptosome, where it interacts with APAF-1 and acts upon the pro- (inactive) form of an initiator (i.e., caspase-9), resulting in activated caspase-9, which then activates the inactive forms of downstream effector caspase molecules (e.g., caspase-3). Once activated, the effector caspases degrade cellular membrane lipids, proteins, and nuclear DNA, which results in the apoptotic cell death of the affected sensory cell. The Bcl-2 family of pro- and anti-apoptotic molecules all act to either inhibit (anti-) or initiate (pro-) pore formation in the mitochondrial membranes. It is the formation of pores in the mitochondrial membrane that releases the cytochrome c from the mitochrondrion into the cytoplasm. Bcl-2 and* Bcl_{xl} *act to inhibit pore formation (anti-apoptotic), and BIM (pro-apoptotic) acts to inhibit the mitochondrial membrane stabilizing action of Bcl-2 and* Bcl_{xl}*. BAX acts directly (pro-apoptotic) to form pores in the mitochondrial membranes.*

commitment to apoptosis. This degradation of nuclear DNA by the caspases and other degradation enzymes (e.g., poly-ARP-ribose-polymerase [PARP] activated by caspase also cause free ends of the degraded DNA to label with the terminal uridine nick translation end-labeling (TUNEL) technique used to identify cells that are in the process of apoptosis. It is this damage to cellular proteins, lipid membranes, and nuclear DNA that is downstream of the release of cytochrome c from the ROS-damaged mitochondria that causes an irreversible commitment of oxidative stress–damaged sensory cells to elimination via apoptosis. Because of all of the cell death events that are activated by cytoplasm localized cytochrome c, the release of this respiratory chain enzyme from the ROS-damaged mitochondria into the cytoplasm is considered to be a significant step in a damaged cell's progress toward a commitment to a cell death (apoptosis) program. Members of the B cell lymphoma molecule two (Bcl-2) family of anti- and pro-apoptotic molecules play important roles in the mitochondrial release of cytochrome c into the cytoplasm of a ROS-damaged cell within the inner ear. It is the balance between the pro- and anti-apoptotic Bcl-2 molecules that either protects against or initiates the depolarization of mitochondrial membranes, formation of membrane pores, and release of cytochrome-c from the damaged mitochondria. BAX and BIM are pro-apoptotic members of the Bcl-2 family that act to enhance pore formation and to inhibit the anti-apoptotic protective effects of Bcl-2 and Bcl_{xl}, members of the Bcl-2 family. BAX directly forms pores in the mitochondrial membranes, and BIM inhibits the anti-pore forming protection of Bcl_2 and Bcl_{xl}, Bcl_2 and Bcl_{xl}, anti-apoptotic factors, act to stabilize mitochondrial membranes and prevent pore formation, preventing the release of cyotchrome c from ROS-damaged mitochondria and the downstream effects of procaspase activation to its active form as a caspase. Thus, mitochondria may act as both a source of ROS and also as the cell executioner in the face of excessive oxidative stress.

Glutamate Excitotoxicity

Glutamate is a major neurotransmitter between inner hair cells and afferent cochlear nerve endings of the type 1 spiral ganglion neurons. It has been postulated that exposure to excessive sound stimulation causes excessive synaptic glutamate concentrations to develop, which lead to overstimulation of the glutaminergic receptors, invoking metabolic cascades, which, in turn, lead to cell injury and death. Some of the potentially harmful cascades set in action by glutamate excitotoxicity may include increases in intracellular calcium with the activation of calcium-dependent calmodulin, the subsequent activation of nitric oxide synthase (NOS), resulting in excessive production of nitric oxide (NO), and the generation of related free radicals such as peroxynitrite. Other glutamate-induced processes may include activation of protein kinases; phospholipase A_2; proteases, such as calpain; and xanthine oxidase, with the subsequent generation of superoxide anions, in addition to mitochondrial injury. Looking to central nervous system glutamate excitotoxicity as an example, glutamate excitotoxicity can be divided into an early phase (up to 30 minutes) and a late phase (3–24 hours) of ROS production consequent to excessive glutamate production. The latter phase of glutamate-induced ROS production occurs as a self-propagating process in which damaged mitochondria become the source of both additional ROS production and further damage to the affected cell.

Evidence for the involvement of glutaminergic synapses in the pathogenesis of NIHL includes the fact that glutamate agonists infused into the inner ear can create an injury to the organ of Corti that is similar in histology to a lesion caused by acoustic overexposure. Also, a variety of glutamate antagonists can partially prevent hearing loss in experiments with animal models of NIHL. Organ of Corti damage produced through noise-induced release of excessive glutamate is thought to occur through a variety of different mechanisms.

At the level of the inner hair cell dendrites, excessive glutamate release is associated with a massive influx of ions and water into these dendritic terminals, leading to loss of the dendritic process. Pujol and colleagues have shown that this glutamate-initiated dendritic damage is at least partially reversible, and that the dendritic processes can be regenerated and form new synapses with the hair cells. Excessive glutamate release is also thought to cause an increase in the activity of NOS. The resulting overproduction of NO, in turn, can lead to the production of several free radicals, such as peroxynitrite, that can directly damage cellular structures. At the level of the OHC, chronic excessive glutamate exposure can lead to intracellular GSH depletion, excessive damage, and then to OHC loss. This may occur through inhibition of the glutamate-cystine antiporter that is responsible for exporting glutamate from cells and importing cystine, which is then converted to a key building block molecule needed for intracellular GSH synthesis. Loss of the GSH ultimately leads to the cell's demise because this protective molecule is key in the body's defense against oxidative stress. As discussed in the next section, other mechanisms may be responsible for noise-induced GSH depletion. These concepts of noise exposure–generated oxidative stress are summarized for the inner hair cells (IHCs) in **Fig. 32–4A** and OHCs in **Fig. 32–4B**.

Glutathione Depletion

NIHL may, in part, be considered to result from the consequences of a temporary reduction in cochlear GSH in the face of oxidative stress. Because GSH is a key antioxidative stress metabolite, its depletion is a consequence of prolonged exposure to oxidative stress; however, its depletion also exacerbates the imbalance in the homeostasis of cochlear tissues caused by the ongoing oxidative injury. Thus GSH depletion is both a consequence and a cause of oxidative cell injury. GSH is one of the key cellular antioxidant molecules present within eukaryotic cells. It is a tripeptide composed of the amino acids glutamate, cysteine, and glycine. The rate-limiting enzyme for its synthesis is γ-glutamylcysteine synthetase, which is regulated by feedback inhibition by its end product, GSH. GSH acts as a free radical scavenger, as well as a detoxification molecule against hydrogen peroxides and other peroxides. GSH keeps vitamins C and E in their reduced active states, maintains the thiol moieties on proteins and peptides in a reduced state, and detoxifies xenobiotics enzymatically through the GSH-transferase family of enzymes or by a nonenzymatic action by forming conjugates. Oxidized (inactive) GSH (i.e., GSSG) is returned to the reduced (active) GSH state through GSSG reductase, and this reaction requires a molecule of the reduced form of nicotinamide-adenine dinucleotide phosphate (NADPH). Methods for increasing

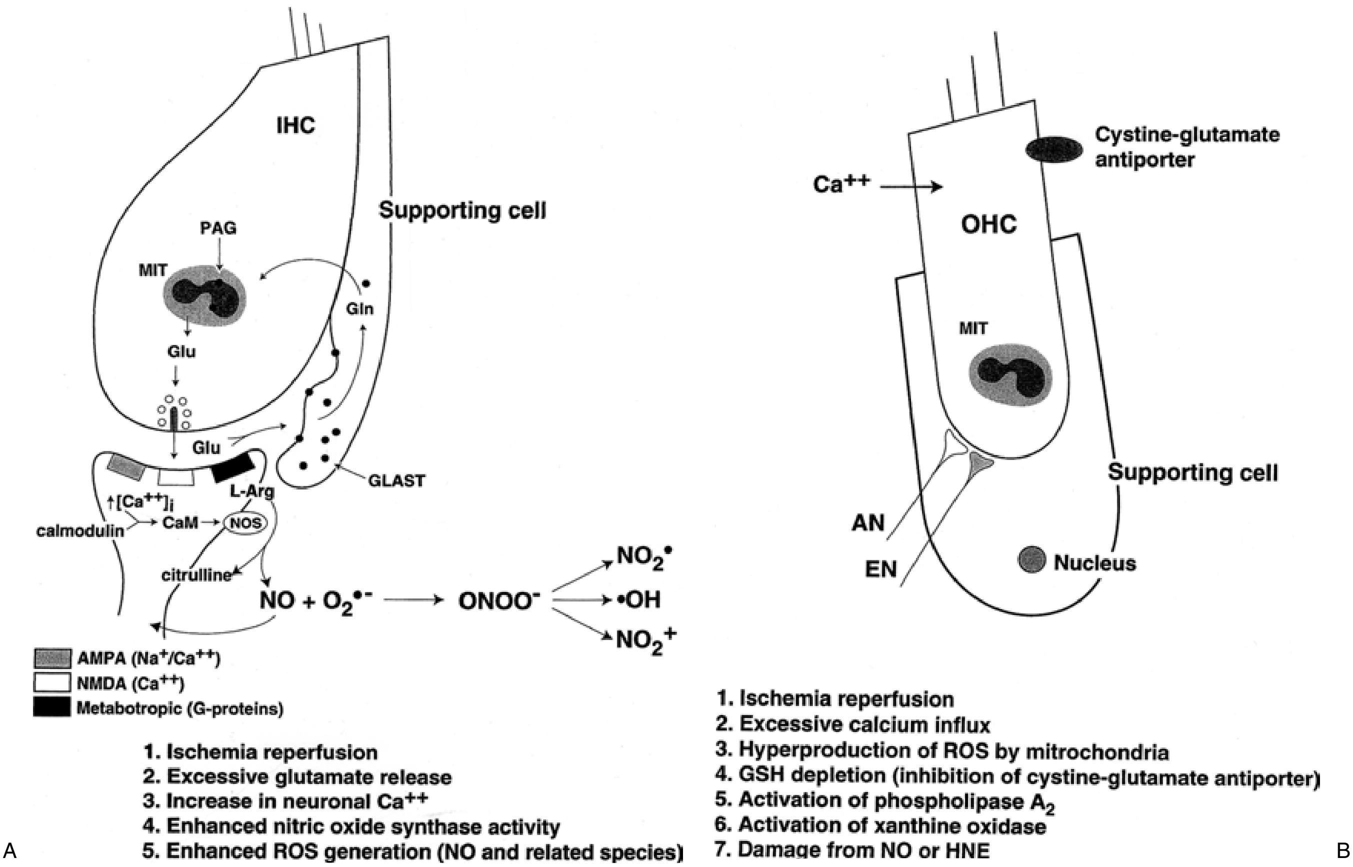


Figure 32–4 *Postulated generators of oxidative stress in the organ of Corti.* ***(A)*** *Inner hair cell (IHC): Excessive release of glutamate (Glu) from overstimulated inner hair cells activates (Alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)-or N-methyl-D-aspartate (NMDA)-s type afferent neuronal postsynaptic receptors. This leads to a massive influx of calcium and other ions into the dendrite, leading to the activation of calmodulin. Nitric oxide synthase (NOS) is then activated to produce nitric oxide (NO) from L-arginine (L-Arg) in conjunction with a variety of other reactive oxygen species. Ordinarily, after being transported into a supporting cell by the glutamate transporter (GLAST), Glu may be metabolized to glutamine (Gln) in a supporting cell by glutamine synthetase and recycled back into the inner hair cell to be resynthesized into Glu by phosphate-activated glutaminase (PAG) located in the mitochondria (MIT). However, the efflux of Glu may be in excess of what the supporting cell GLAST can metabolize. Oxygen deprivation secondary to ischemia reperfusion may play a role in several of the depicted changes. (Modified from Nordang L. Cestreicher E, Arnold W, et al. Glutamate is the afferent neuro-transmitter in the human cycle. et al. Acta Otolaryngol 2000;120:359–362.)* ***(B)*** *Outer hair cell (OHC): OHCs are separated from IHCs by the tunnel of Corti and have reduced numbers of glutaminergic synapses compared with those present at the bases of the IHCs. The cystine-glutamate antiporter is shown. Ischemia-reperfusion, increases in intracellular calcium, and hyperproduction of ROS by OHC mitochondria (MIT) may play a role in subjecting OHCs to oxidative stress. Inhibition of the cystine-glutamate antiporter may be a source of oxidative stress unique to the OHC. 4-hydroxy 2,3-nonenal (HNE) may act as a relatively stable toxic molecule generated by oxidative-stress and be able to diffuse between cells. AN, afferent neuron; EN, efferent neuron; NO, nitric oxide.*

cellular GSH include the administration of an esterified form of GSH, methionine; L-Nacetylcysteine (NAC); L-2-oxothiazolidine-4-carboxylate (OTC); or γ-glutamylcysteine.

Glutathionine is not cell permeable, and to have a protective effect against ROS formation and ROS damage, GSH must be inside the cell. The esterified forms of GSH (i.e., monoethyl ester and diethyl ester) are cell permeable and are highly effective as a treatment to prevent ROS damage within an oxidative stressed cell. The monoethyl ester is more stable than the diethyl ester of GSH and therefore a better protective molecule for long-term administration. Esters of GSH when placed on the round window membrane (RWM) can protect the cochlea from impulse and continuous noise damage. GSH precursor molecules, OTC, and NAC (a source of the GSH precursor molecule cysteine) have been shown to reduce noise-induced threshold shifts when administered systemically prior to either continuous exposure to noise or exposure to an impulse noise.

Increases in Intracellular Calcium

One theory of noise-induced cochlear injury holds that acoustic overexposure leads to an excessive influx of calcium ions into the hair cells, leading to injury and death of the affected cells. Ionic calcium plays an important, but as yet incompletely defined, role in cochlear hair cell physiology, including regulation of transduction, neurotransmitter release, OHC slow motility, and OHC adaptation. Calcium ions normally enter hair cells through L-type calcium channels and voltage-sensitive calcium channels.

During acoustic overexposure, harmful elevations in intracellular calcium may occur as a result of ROS-induced or micromechanically induced damage to hair cell plasma membranes and ROS-induced injury to calcium regulatory proteins, and organelles, such as mitochondria and mechanical stimulation. Increases in cellular calcium levels can then induce a cascade of events, including activation of phospholipases (damage cell membranes), protein kinase C (disrupt microtubules), proteases such as calpain (alter membrane permeability), and endonucleases (impair protein synthesis). Activation of these processes may lead to cell injury and/or death of the cell.

AMELIORATING NOISE-INDUCED HEARING LOSS THROUGH PHARMACOLOGICAL THERAPIES

Given the theories outlined in the preceding section, it follows that approaches to prevent or reverse acute NIHL should be designed based on known mechanisms of pathogenesis. Hence, approaches to mitigate mitochondrial injury, reduce glutamate excitotoxicity, replenish GSH, and reduce harmful intracellular calcium fluxes have all met with some measure of success (novel strategies are presented and discussed in the following section).

Protecting/Restorating Mitochondria

Seidman and Van De Water (2003) have shown that a mitochondrial metabolite known as acetyl-L-carnitine (ALCAR) was effective in reducing age-related hearing loss in a rat model of presbycusis. Age-related hearing loss is also thought to be related to the accumulation of cochlear damage from chronic oxidative stress. We have shown that ALCAR is also effective at reducing NIHL in a chinchilla model when given before and after intense, continuous noise exposure (**Fig. 32–5**). ALCAR supplies acetyl moieties and carnitine. These molecules enhance mitochondrial energy production and restore a key mitochondrial molecule known as cardiolipin, while restoring mitochondrial membrane integrity. This restores mitochondrial efficiency with a consequent reduction of mitochondrial free radical formation, preventing mitochondrial-induced apoptosis.

Prevention of the release of cytochrome c from damaged mitochondria is also an effective therapeutic approach to reduce NIHL that results from excessive noise exposure. Cytochrome c release from damaged mitochondria can be prevented by preventing pore formation (e.g., overexpression of Bcl-2 or Bcl_{xl} via a gene therapy vector) or interruption of the signaling cascade of JNK/c-Jun cell death signaling pathway (e.g., c-Jun antisense therapy) that damages the mitochondria of oxidatively stressed auditory sensory cells.

Reducing Glutamate Excitotoxicity

Several strategies are possible for reducing noise-induced glutamate toxicity, including countering the damaging ionic fluxes nonspecifically through magnesium supplementation, or more specifically, through the application of an antagonism of glutamate receptor–associated ionic channels. Specific glutamate receptor antagonism can be accomplished, and NOS can be inhibited. In addition, treatment with esterified GSH can ameliorate the toxicity associated with excessive glutamate release.

Carbamathione, a glutamate receptor antagonist, given systemically to chinchillas beginning prior to 6 hours of exposure to a 105 dB SPL 4 kHz octave band noise, reduced both hearing loss and hair cell loss (**Fig. 32–5**). The decrease in NMDA receptor binding by glutamate is thought to be due to a modification of the redox modulator site on the receptor. Many glutamate antagonists available to date are associated with undesirable side effects due to the excessive activity of some of these blocking agents. Carbamathione [*S*-(*N,N*-diethylcarbamoyl) glutathione, also known as DETC-GS], unlike classical glutamate antagonists that yield complete inhibition at the level of interaction with the receptor (e.g., CGS 19755) or directly at receptor-linked, calcium ion channels (e.g., phencyclidine or MK 801), is thought to impart its inhibitory effects via interaction with the redox modulatory site of the NMDA receptor. The latter type of interaction produces partial glutamate antagonism, is selective for the NMDA subtype of glutamate receptor, and should produce fewer unwanted side effects.

Reducing GSH Toxicity

Maintaining, enhancing, and restoring cochlear GSH levels have several potential advantages as a treatment strategy to reduce NIHL. First of all, GSH is a key intracellular antioxidant and inhibitor of stress-induced

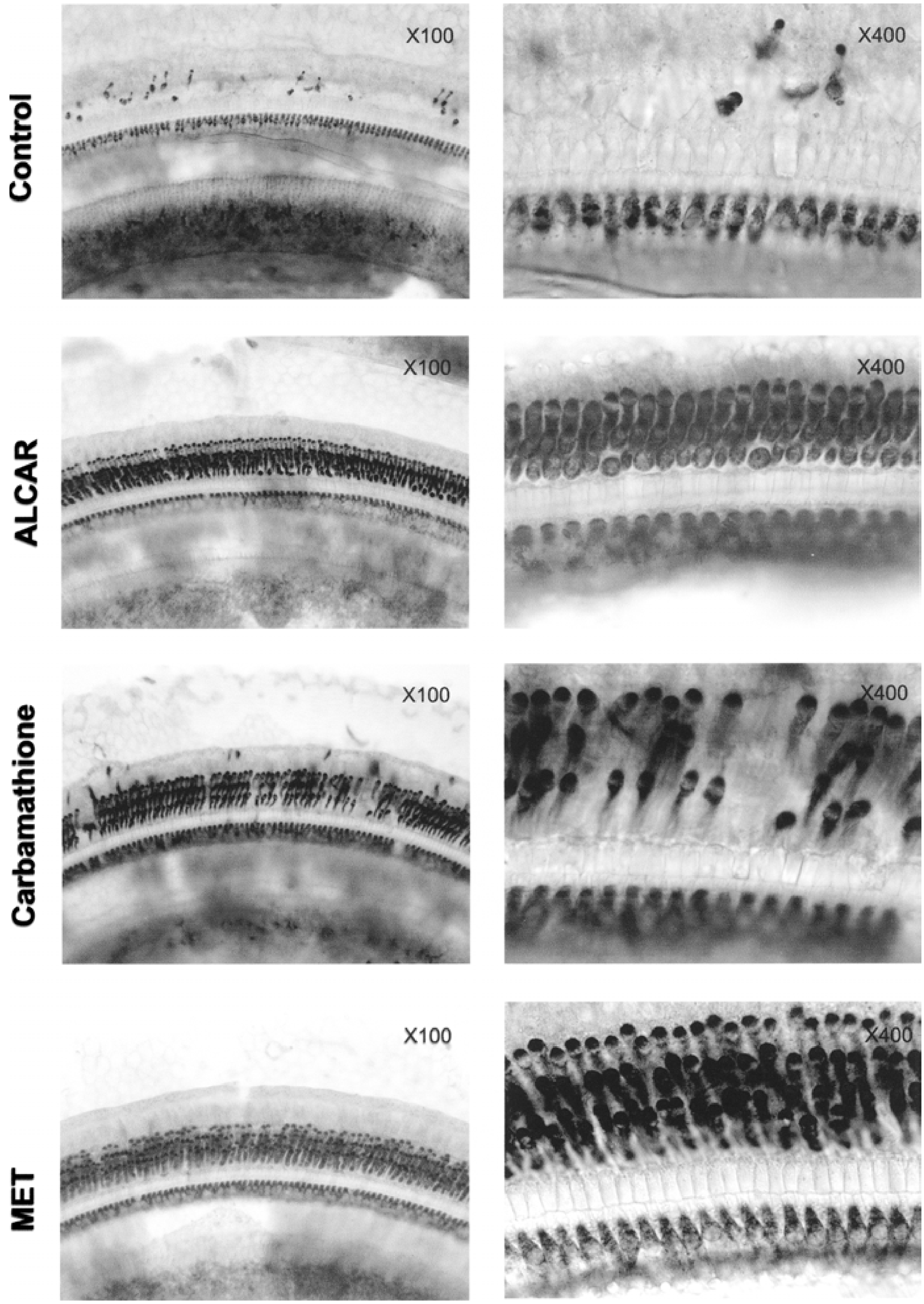


Figure 32–5 *Treatment of noise-exposed animals with either acetyl-L-carnitine (ALCAR) or methionine (MET) prevents loss of hearing and auditory hair cells. Photomicrographs are cochlear surface preparations, with the hair cells stained with sodium succinate histochemistry. Micrographs on the left are low-power (100×) and those on the right are high-power (400×) images of the 6 kHz regions of cochleas. All photomicrographs were taken from noise-exposed animals treated with either saline (control) or ALCAR, carbamathione, or MET, depicted in rows from top to bottom. The single row of inner hair cells (IHCs) is oriented toward the bottom, and the three rows of outer hair cells (OHCs) are toward the top of each micrograph. The control animal demonstrates an almost total loss of viable (stained) OHCs, with a scattered loss of IHCs in the sound trauma lesion site. In contrast, micrographs from MET- and ALCAR-treated noise-exposed animals demonstrate a single row of intact IHCs and three complete rows of viable OHCs, similar to the pattern of these sensory cells in non-noise-exposed cochleas (not shown). Carbamathione-treated animals demonstrated only partial protection of the cochleas from sound trauma with an intermediate level of OHC losses.*

apoptosis. GSH is an effective detoxifying molecule for lethal lipoperoxides, such as HNE, that may be produced as a consequence of excessive noise. Additionally, GSH can counter the harmful effects of glutamate excitotoxicity, mitochondrial injury, and excessive intracellular calcium fluxes. Thus GSH levels seem to be important in ameliorating the damage caused by all of the pathological mechanisms described for noise-induced, cochlear oxidative stress mentioned in this review. GSH is rapidly metabolized on its first pass through the liver and is poorly transported into most cells; it therefore has limited bioavailability.

GSH is not cell permeable and must be present within a cell to provide effective otoprotection; therefore, treatment with unmodified GSH is ineffective as a therapy. Fortunately, intracellular GSH levels may be enhanced through the use of a variety of GSH or cysteine precursors and cell-permeable forms (esters) of GSH. These agents, such as NAC, methionine, GSH esters (GSHe), and thiazolidine-related drugs, such as 2-oxothiazolidine-4-carboxylate, have in common that they are readily transported into cells and serve as intracellular sources of cysteine, which can be used by the cell to produce GSH, or in the case of GSH esters have a direct action. All have been shown to reduce

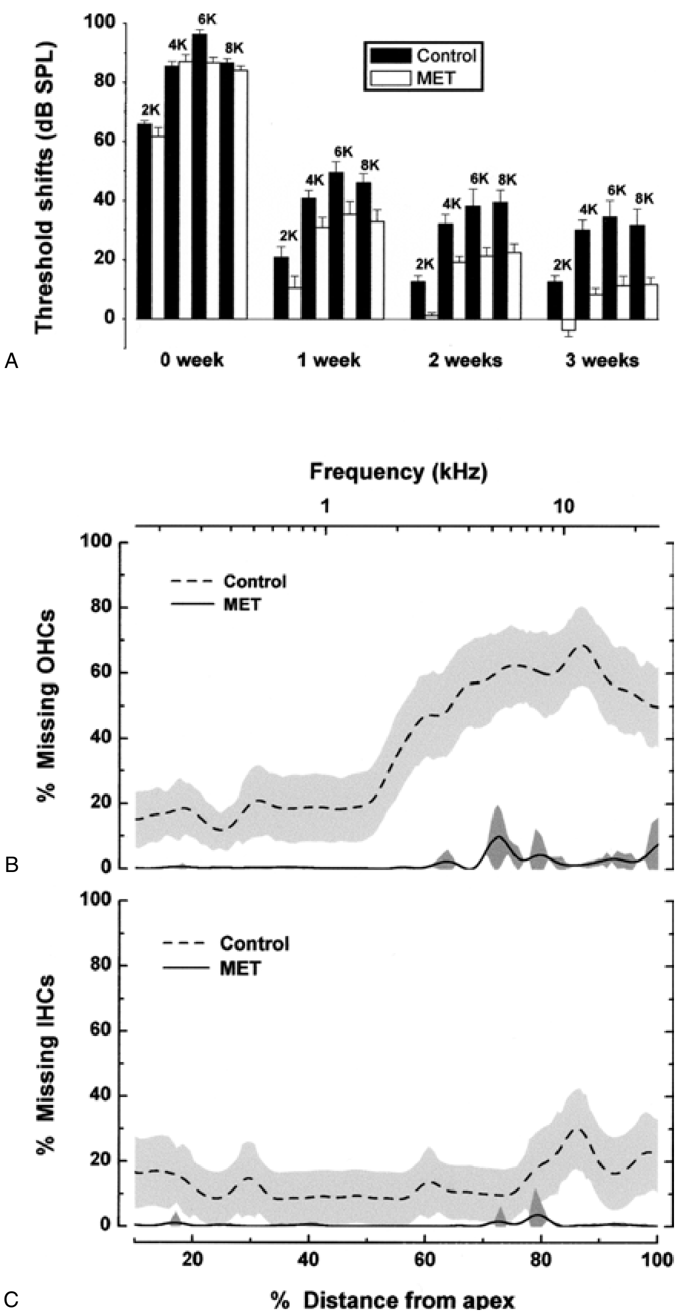


Figure 32–6 *Methionine (MET) treatment protects hearing from noise exposure–generated hearing loss and loss of auditory hair cells.* ***(A)*** *Auditory threshold shifts for saline-treated and MET-treated animals following exposure to a damaging level of noise. Means for auditory threshold shifts (dB SPL) plotted as a function of treatment group (saline control-noise and MET treatment-noise), time [0 (1 hour), or 1, 2, 3 weeks post-noise exposure], and by threshold test frequency for 2, 4, 6, and 8 kHz. Initial threshold shifts (week 0) ranged from ~62 to 87 dB for the MET-noise group, which were statistically similar to the saline-treated noise-exposed group ($p > .05$), except for 6 kHz, where the mean threshold was significantly less than that of controls ($p < .05$). An overall treatment effect for the MET-treated group was protective when compared with the saline-noise group ($p < .001$) for all test frequencies beginning at week 1. Error bars are ± standard error of means (SEM) Sample (n) size is 12 for all groups (12 ears, six animals).* ***(B)*** *Outer hair cell (OHC) cytocochleogram results. Depicted are mean values (continuous line) and standard error of the mean (shaded area) cytocochleograms for OHCs MET-pretreated noise-exposed cochleas (solid line) and saline-treated noise-exposed cochleas (dotted line), respectively. The y-axis depicts mean percent missing OHCs. The lower x-axis represents percent distance from the cochlear apex, and the upper x-axis depicts the associated frequency range of the cochleas in kHz. A very small loss of OHCs occurred in the low-dose MET-protected cochleas (<10%), whereas substantial OHC losses occurred in the saline-control noise-exposed group (average of ~60% for the 4–10 kHz region). These differences were significant ($p < .001$). Error bars are ± standard error of means SEM. Sample (n) size is 12 for all groups (12 ears, six animals).* ***(C)*** *Inner hair cell (IHC) cytocochleogram results. Illustrated are the mean values of IHC cytocochleograms with missing IHC percentages on the y-axis as a function of the measured percent distance from the cochlear apex. The associated frequency region of the cochlea in kHz is also plotted on the upper x-axis, and the percent distance from the cochlear apex is depicted on the lower x-axis. MET treatment (solid line) afforded significant protection of IHCs, as seen by a reduction to 5% or less of IHC loss, with MET treatment versus over 20% in the saline-treated animals ($p < .05$). Error bars are ± standard error of means SEM. Sample (n) size was 12 for all groups (12 ears, six animals).*

NIHL in laboratory studies. Results for protection against NIHL for methionine treatment are shown in **Figs. 32–6A, B,** and **C**.

Ameliorating the Effects of Increases in Intracellular Calcium

Approaches to reduce NIHL through the use of calcium channel blockers have been published in both the basic science and the clinical literature. The calcium channel blocker diltiazem was utilized pre- and postoperatively in a prospective, randomized, double-blind study (Heinrich et al., 1999; Maurer et al., 1995) as an approach to reduce noise trauma associated with otologic surgery. There was a tendency toward reducing noise-induced permanent threshold shifts in those patients given diltiazem, but this effect on NIHL did not reach statistical significance. Additional work in an experimental model has shown that diltiazem may modulate precipitable calcium in the hair cells in guinea pigs after noise exposure, as well as reduce damage to the organ of Corti caused by a sudden impulse noise. However, other studies in mice and gerbils found no protective effect from noise with either diltiazem or another calcium channel blocker, nimodipine. With regard to diltiazem, it may not achieve an adequate drug level in the inner ear due to issues related to its ability to cross the blood-cochlear barrier. At present, calcium channel blockers have not been found to be consistently effective in reducing NIHL in experimental models of sound trauma.

A variety of approaches can be imagined to reduce NIHL pharmacologically (see summary in **Fig. 32–7A,B**). Several GSH-restoring compounds are in clinical use at this time and are orally well tolerated with few side effects. These agents can be given in tablet form prior to loud noise exposure in the workplace or in a recreational setting. Recent evidence shows that NAC and MET, two molecules used for intracellular GSH synthesis, can reduce NIHL substantially when given as a treatment agent shortly after noise exposure. As previously mentioned, GSH has multiple positive effects, including protecting and restoring mitochondrial function, reducing the effects of glutamate excitotoxicity, reducing the injury associated with excessive intracellular calcium levels, and preventing stress-induced apoptosis. Because of these many positive effects, the strategy to reduce acute NIHL by augmenting inner ear GSH levels or direct treatment with an esterified form of GSH is quite appealing. An additional strategy of clinical utility may be to treat with the mitochondrial metabolite ALCAR. ALCAR has been shown to be effective in reducing NIHL when given around the time of the acute noise

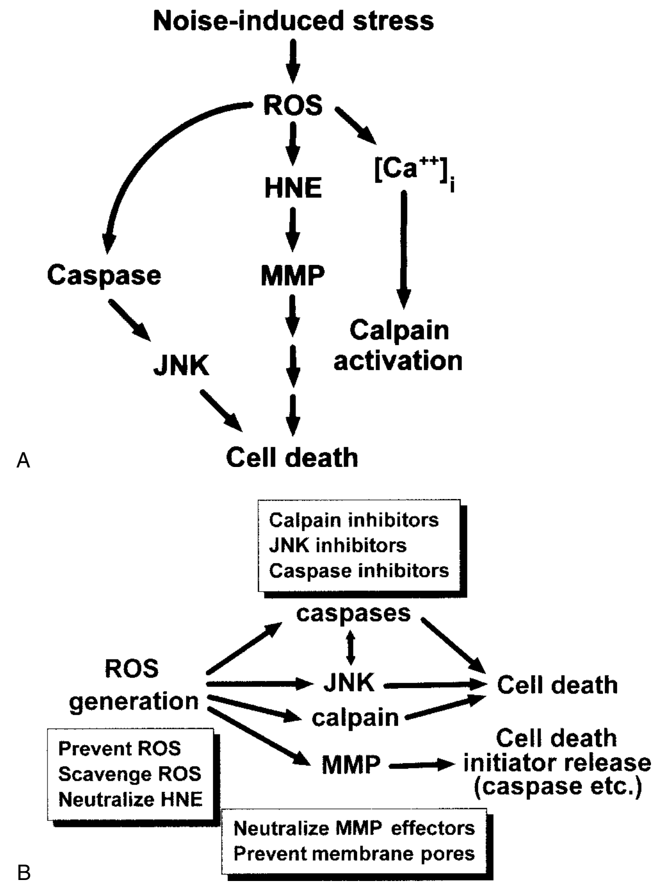


Figure 32–7 *Noise-damage activation of cell death (apoptosis) within the cochlea.* ***(A)*** *Cell death pathways. An outline of how noise-induced stress can lead to cell death in the cochlea. Multiple pathways are likely to be involved. Upstream events include the generation of reactive oxygen species (ROS) and perhaps a pathological increase in intracellular calcium. 4-hydroxy 2,3-nonenal (HNE) is likely formed downstream of the noise-damage-induced burst of ROS production. Increases in intracellular calcium also may occur at several points due to ROS-induced cell injury, including mitochondrial injury. Mitochondrial membrane permeablization (MMP) is a central event in the cell death process. Caspases may be activated either following MMP or independent of MMP. Caspases may activate the c-Jun-N-terminal kinase and c-Jun (JNK/c-Jun) cell death pathway downstream of MMP or be activated independently and upstream of MMP. Caspase activation also can occur downstream of (JNK/c-Jun) activation (see* ***Fig. 32–3****). Upstream activation is depicted in this figure. Calpain activation, caspase activation, JNK activation, and MMP can all lead to cell death by interrelated and also by independent pathways.* ***(B)*** *Potential points of intervention for the treatment of NIHL and apoptosis of auditory sensory cells. Upstream points of intervention include the use of antioxidant and free radical scavengers to prevent the formation of ROS, to scavenge ROS, and to neutralize HNE. More downstream interventions include the use of inhibitors of calpain, JNK, and caspases. Another downstream strategy would be to neutralize MMP effectors and/or to prevent mitochondrial membrane pore formation.*

exposure. High dosage levels of ALCAR are approved for clinical application and have been used for a year or longer to treat chronic neurodegenerative diseases and diabetes. The use of glutamate antagonists shows promise; however, clinical side effects with these agents may limit their usefulness. Experimental evidence to date would suggest that calcium channel blockers are less likely to be effective clinically than the other agents already listed and discussed in this chapter.

Another set of interesting treatment strategies might be to treat the noise-injured cochlea with drugs that block programmed cell death pathways. While GSHe, MET, and perhaps ALCAR can all fulfill this role, other more specific apoptotic pathway inhibitors might be employed. For example, leupeptin, a calpain inhibitor, has been found to be somewhat effective in reducing noise injury to the cochlea; there are both pancaspase inhibitors and specific inhibitors of caspases that have yet to be tested as a protective therapy against NIHL. Investigators have published success in reducing NIHL in animal models by inhibiting steps in the mitogen activated protein kinase (MAPK) signaling pathway of apoptosis and more specifically, mixed lineage kinases (MLKs) that lead to the activation of the JNK/c-Jun segment of this pathway and also a peptide inhibitor that directly blocks the activation of the JNK molecule.

Finally, a future strategy for revising long-standing NIHL might involve regeneration of cochlear sensory hair cells coupled with the reestablishment of functional synapses. Birds have been shown to form new auditory hair cells through cell division of support cells followed by differentiation of new hair cells and support cells and/or transdifferentiation of existing support cells as sources of replacement hair cells. The newly formed replacement hair cells acquire new synaptic connections and restore auditory function. Mammals have not been shown to form new hair cells after injury in mature animals. This may be explained in part by the presence of cell proliferation inhibitors found to be active in the cells of the mature mammalian cochlea. Nevertheless, in the future it may be possible to override this block on proliferative regeneration through the use of trophic factors alone or in combination with inhibitors of cell proliferation inhibitors. Genes important for the genesis of hair cells in mammals, including *Notch*, *Math 1, Brn3.1* and many others, are just being discovered. Therefore, it may become possible to manipulate the expression of these genes or to deliver these genes to cochlear epithelial tissues to effect a hair cell regeneration response with return of function in humans. An initial study of *Math 1* gene therapy in a mammal suggests that this may be possible. This approach will require much additional research, and the possibility of the renewal of cochlear hair cells in arrears of damage and hair cell loss, although distant, seems real.

CONCLUSION

NIHL remains an important 21st century otolaryngology health problem responsible for permanent hearing loss in some 10 million individuals in the United States alone.

Current hearing conservation measures involving engineering, avoidance, and the use of HPDs remain important but have their limitations. A future comprehensive hearing conservation program may include pharmacological treatment to prevent and/or reverse acute NIHL. The "oxidative stress hypothesis" of noise-induced cochlear injury and loss of auditory sensory cells suggests a variety of pharmacological treatment strategies to ameliorate NIHL. These strategies include the oral administration of several different compounds before or after noise exposure. These compounds include agents to replenish cochlear GSH (e.g., NAC), esterified GSHe, mitochondrial protectants (e.g., ALCAR), glutamate toxicity antagonists, or calcium channel blockers, or agents to prevent noise-induced cochlear vasoconstriction. Additionally, specific programmed cell death pathway inhibitors (e.g., round window membrane application of c-Jun antisense oligonucleotide therapy or D-JNKI-1 treatment) may prove to be useful clinically. Finally, the ability to initiate auditory sensory cell regeneration through gene therapy looms in the future as a challenge, an exciting possibility, and a worthwhile goal to restore not only NIHL but also sensorineural hearing loss from other causes.

NOTE

The views expressed herein are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, or the United States government.

Suggested Readings

Axelsson A, Dengerink H. The effects of noise on histological measures of the cochlear vasculature and red blood cells: a review. Hear Res 1987;31:183–191

Heinrich UR, Maurer J, Mann W. Ultrastructural evidence for protection of the outer hair cells of the inner ear during intense noise exposure by application of the organic calcium channel blocker diltiazem. ORL J Otorhinolaryngol Relat Spec. 1999;61:321–327

Henderson D, Hamernik R. Biologic bases of noise-induced hearing loss. Occup Med 1995;10:513–534

Kopke RD, Allen KA, Henderson D, et al. A radical demise: toxins and trauma share common pathways in hair cell death. Ann NY Acad Sci 1999;3:171–191

Kopke RD, Coleman JKM, Huang X, et al. Novel strategies to prevent and reverse noise-induced hearing loss. In: Henderson D, Prasher D, Kopke R, Salvi R, Hamernik R, eds. Noise-Induced Hearing Loss: Basic Mechanisms, Prevention and Control. London: Noise Research Network Publications; 2001: 231–253

Kopke RD, Coleman JKM, Liu K, Campbell K, Riffenburgh RH. Enhancing intrinsic cochlear stress defenses to reduce noise-induced hearing loss. Laryngoscope 2002;112:1515–1532

Maurer J, Riechelmann H, Amedee RG, Mann WJ. Diltiazem for prevention of acoustical trauma during otologic surgery. ORL J Otorhinolaryngol Relat Spec. 1995;57:319–324

Puel JL, Ruel J, Guitton M, Pujol R. The inner hair cell afferent/efferent synapses revisited: a basis for new therapeutic strategies. Adv Otorhinolaryngol 2002; 59:124–130

Seidman M, Van De Water TR. Pharmacological manipulation of the inner ear [review]. ENT Journal 2003;82:276–288

Seidman MD, Van De Water TR. Pharmacologic manipulation of the labyrinth with novel and traditional agents delivered to the inner ear. Ear Nose Throat J 2003; 82:276–300

Van De Water TR, Lallemend F, Eshraghi AA, et al. Caspases, the enemy within: their role in oxidative stress-induced apoptosis of inner ear sensory cells [review]. Otol Neurotol 2004

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716

1. Limitations of mechanical personal hearing protection devices include
   - A. Noise energy exceeds protection afforded by the device.
   - B. Transmission of damaging acoustic energy through the skull
   - C. Improper fit of the device
   - D. Limitations in protection from different frequencies of noise
   - E. All of the above
   - F. A and C only

2. The oxidative stress hypothesis of noise-induced hearing loss
   - A. Is supported by the finding of noise-induced changes in cochlear glutathione (GSH), lipid peroxidation products, antioxidant enzyme activity, and free radical concentrations
   - B. Predicts that oxidative injury is the major cause of cochlear damage regardless of the degree of sound pressure level exposure
   - C. Is consistent with the finding that agents that increase cellular GSH decrease noise-induced cochlear damage
   - D. Is contradicted by the finding of the activation of programmed cell death pathways in the cochlea after intense noise exposure
   - E. A and C only
   - F. B and D only
   - G. A through D

3. Involvement of mitochondria in noise-induced cochlear injury can be described by which of the following?
   - A. Mitochondria are a major source of oxidative stress.
   - B. Mitochondria play an important role in the initiation of cell death processes after acoustic exposure.
   - C. Mitochondria are insensitive to glutamate levels induced by noise.
   - D. Mitochondria furnish their own GSH and are resistant to GSH depletion.
   - E. All of the above
   - F. A and B only

4. The time from acute narrow-band noise exposure to the loss of the majority of outer hair cells and establishment of a permanent hearing loss is best measured in
   - A. Minutes
   - B. Hours
   - C. Days
   - D. Weeks
   - E. Months

Chapter 33
VESTIBULAR SYSTEM PHYSIOLOGY

John Carey

The vestibular system can be thought of as a system that senses and controls motion. Its functions begin with the detection of head position and motion by the vestibular end-organs. The vestibular hair cells transduce the mechanical stimuli into neural signals and convey them to the brainstem. The brainstem processes the signals to generate secondary neuron messages appropriate to specific reflex tasks. Here we examine in particular two of the brainstem's tasks: integration and velocity storage. These tasks have important effects on our tests of vestibular function. The brainstem then distributes its secondary vestibular neuron signals to other areas of the central nervous system (CNS) to generate vestibular reflexes and sensation. Specifically, neuronal signals are sent to the oculomotor nuclei to generate the vestibulo-ocular reflex (VOR), which stabilizes our gaze (eye position in space). Other neurons go to the cervical spinal motor neurons to generate the vestibulocolic reflex (VCR), and to the lower spinal motor neurons to generate the vestibulospinal reflexes (VSRs). These reflexes stabilize both our posture and our gait. There are cortical pathways to provide us with a conscious sense of motion. Autonomic pathways have long been known, but only recently have we begun to appreciate their function. Vestibular input, particularly otolith information about our posture with respect to gravity, is used here to adjust hemodynamic parameters to maintain cerebral perfusion. Finally, vestibular input to the cerebellum is critical to coordination of the vestibular motor reflexes and to the ability to adapt those reflexes when changes occur, such as injury to a vestibular end-organ or alteration in vision (e.g., a new pair of glasses).

THE VESTIBULAR PERIPHERY

The three semicircular canals with their associated ducts are sensors of angular head acceleration. Each begins and ends in the utricle. The business end of the semicircular duct within the canal is the ampulla, a swelling near the utricle that houses the sensory receptor apparatus. The horizontal (or lateral) canal is actually pitched up ~20 degrees from the horizontal when the head is upright. There are two vertical canals: the posterior (or inferior) and the anterior (or superior). These three semicircular canals lie at almost right angles to each other, so that any angular motion in three-dimensional space can be detected.

There are two otolith organs (maculae) located in the utricle and saccule. These are sensors of linear acceleration, whether it is either a to-and-fro movement of the head or gravity acting on the stationary head. The macula of the utricle lies approximately in the horizontal plane, and the macula of the saccule lies approximately in the vertical, parasagittal plane. The superior vestibular nerve supplies the receptor epithelium of the horizontal and anterior canal cristae, the macula of the utricle, and part of the saccular macula. The inferior vestibular nerve supplies afferent innervation to the receptor epithelium of the posterior canal crista and most of the macula of the saccule.

ADEQUATE STIMULI FOR VESTIBULAR RECEPTOR EPITHELIUM

All vestibular hair cells sense shearing forces. When the hair bundle composed of stereocilia is deflected toward the kinocilium, the hair cell is excited. Its receptor potential goes up, it releases more neurotransmitter, and the firing rate of the underlying afferent nerve increases from its baseline-level firing rate, which is already high (~90 spikes per second). Conversely, if the sensory hair bundle is deflected away from the kinocilium, the process is reversed, and the firing rate of the afferent nerve goes down. However, the responses of the vestibular system are not symmetrical. There is a greater increase in firing for a given excitatory deflection than there is a decrease for the same magnitude inhibitory deflection. Incredibly, Flourens and Ewald figured this out in the nineteenth century. By cannulating individual ducts of the semicircular canals in pigeons, they showed that the nystagmus produced by endolymph displacement in one direction was greater than in the other direction. This observation was codified as Ewald's second law.

The advantage of having a high baseline-level firing rate for the vestibular nerve is twofold. First, there is a very low threshold to sensation. The stimulus does not have to reach some high level before the afferents respond. Rather, a very small stimulus will be detected as a modulation in the firing rates of already-active afferent nerves. So, for example, the perceptual threshold to angular acceleration is as low as 0.1 degree/sec^2. Moving at that speed in a swivel chair, it would take an individual almost 90 seconds to go one revolution. Linear accelerations as small as 5×10^{-4} g can be detected. It would take an elevator moving at this acceleration ~40 seconds to travel one floor.

The second advantage of the high baseline-level firing rate of vestibular nerve is that we can sense movement in the off direction. If the nerve were silent at rest, there would be no way of knowing when the hair bundle was being deflected in this "off" direction (i.e., the firing rate cannot become some negative number). Instead, this high baseline-level firing rate gives us a bidirectional response from each ear.

Thus the labyrinth has an incredible dynamic range of head velocities that can be detected. But the range is not infinite, a fact that becomes very important in interpreting vestibular test findings in the presence of unilateral hypofunction. With sufficiently high head accelerations in the off direction, some afferent nerves eventually do reach zero firing, or "cutoff." When this cutoff occurs, the system exhibits what we call "nonlinearity." Up to that point, if we doubled the acceleration in the off direction, we saw that the firing rate on that side would go down by half. Now, if we double the acceleration again, nothing happens to the firing rate; it remains at zero. Thus the linear relationship is lost.

Nonlinearity is not a problem if both sides of the vestibular system are functioning. The excitatory signal from the side toward which the head is turning continues to inform the brainstem about the motion (and can do so up to very high rates, i.e., ~400 spikes per second). If the labyrinth on that side is not functioning, and we have attained a high enough acceleration to silence afferents on the intact, inhibited side, then the brainstem sees no further change in firing from afferents on either side. It therefore loses the ability to detect any movements of higher acceleration.

How do shearing forces actually develop in the vestibular end-organs? In the maculae of the utricle and saccule, the hair cells are arrayed in a horizontal or vertical sheet, respectively. The hair bundles insert into a gelatinous layer called the otoconial membrane. On top of this lie the otoconia or otoliths (i.e., crystals of calcium carbonate). The otoconial membrane is therefore denser than the surrounding endolymph, and tilting the head results in a kind of "slab-avalanche" movement that carries the hair bundles with it. In the semicircular canals, the hair cells are arrayed on a crest, the crista ampullaris, which extends across the ampulla perpendicular to the orientation of the canal. Another gelatinous membrane, the cupula, extends from this crest up to the roof of the ampulla, sealing the canal side off from the utricular side. When relative endolymph motion occurs in the canal, the cupula puckers in the center like a sail, and the attached hair bundles are deflected with it.

DIRECTIONAL SENSITIVITY AND SEMICIRCULAR CANAL PLANES

The exact geometry of hair cell orientation determines the directions of forces to which a given end-organ will respond. Again, the direction of the kinocilium determines the "on" direction for the hair cell. In the semicircular canal cristae all of the kinocilia point in the same direction. In the horizontal canals, they point toward the utricle, and movement of endolymph toward the utricle (utriculopetal) excites the end-organ. This is what occurs in the right ear, for example, when the head turns to the right. In the vertical canals, however, the polarization is reversed. All of the kinocilia point away from the utricle, and utriculofugal flow of endolymph excites the sensory cells of these canals. This is what occurs, for example, in the superior canal when the head is pitched down, and in the posterior canal when the head is pitched up.

One of the important consequences of these arrangements is that canals work in "push-pull" pairs, and the six canals form three canal planes. The horizontal canals (HCs) are paired in the horizontal plane, and a yaw movement that excites one horizontal canal always inhibits the other. The vertical canals pair as follows: the right anterior (RA) canal works in conjunction with the left posterior (LP), forming the RALP plane, which lies ~45 degrees off the sagittal plane; at almost a right angle to this plane, the left anterior (LA) canal pairs with the right posterior (RP), forming the LARP plane. Mixed vertical-torsional head movements stimulate these vertical canals. Remember that each canal is excited by turning the head toward the side of that canal in the plane of that canal. Thus the left HC is excited by leftward rotation in the HC plane. Pitching the head down and rolling it to the right in the RALP plane excites the RA canal and inhibits the LP canal. Pitching the head up and rolling it to the left in the RALP plane reverses the pattern: the LP canal is excited, and the RA inhibited.

The canals drive eye movements in these planes. This is Ewald's first law. A pure stimulation of one of the horizontal canals, for example, will drive the eyes conjugately in the horizontal plane. Likewise, pure stimulation of the left posterior canal, as occurs in benign paroxysmal positional vertigo (BPPV), will drive the eyes in the RALP plane.

The relationship of the two maculae is even more complicated. There is a prominent dividing curve going through the middle of each macula. This plane of separation in each macula is called the striola. Hair cells on opposite sides of the striola have an opposite polarization of the relationship of their stereociliary bundle to the kinocilium. In the macula of the saccule, all of the polarization vectors point away from the striola. In the utricular macula, they all point toward the striola. However, the net result is not a null vector. In fact, in the utricle there is a 3:1 predominance of vectors pointing ipsilaterally. This becomes important in explaining the skew and tilt reactions seen in patients with unilateral labyrinthine hypofunction.

SEMICIRCULAR CANALS AFFERENT PHYSIOLOGY

We now move out of the end-organs and consider the signals carried to the brainstem. One might suppose that the semicircular canals would act like simple tubes of fluid with inertia, and that the fluid, sitting still in the canals as they moved around with the head, would faithfully inform the brain about the acceleration of the head. After all, by Newton's laws, force (F) = mass (m) × acceleration (a) ($F = ma$), so the acceleration of the head should result in a proportional force on the endolymph and the cupula of the ampulla. One of the remarkable things that Steinhausen predicted in the 1930s, however, was that the canals instead would sense velocity. This is because he included in his calculations the very significant elastic restoring force of the gelatinous cupula. This force dominates the system, and the result of his "torsional pendulum" model is that the canals act as mechanical integrators, converting an acceleration signal into a velocity signal. In panel A of **Fig. 33–1**, we see this manifested as we record the firing rate of a horizontal canal afferent. At the top, the stimulus given is a step of acceleration in the "on" direction. Just below that in panel B of **Fig. 33–1** is shown the mathematical integration of this step. It is a ramp, and, by definition, is the velocity of the stimulus. Now look at the firing rate and see that it rises more like a ramp than a step. We say that this afferent is encoding velocity.

What is the significance of this to the VOR? The eye muscles do not need an acceleration and position signals. Rather, their actions can be derived from a velocity signal. Thus the semicircular canals perform the first and positions signal processing necessary to match the input signal characteristics to the reflex output needs. In fact, the integration performed by the semicircular canals is incomplete. As we shall discuss on this section, the brainstem further processes the signals to complete the integration needed for the control of eye movements.

Another manifestation of this integration appears in responses to sinusoidal rotations, which play an important role in vestibular testing. When we differentiate a sine wave, we get the cosine. This has all the same values as the sine wave, but they occur 90 degrees earlier in the cycle. When the two (sine input and cosine output) are plotted together, however, it appears that the output is lagging behind the input by 90 degrees. We say that the response lags the stimulus by 90 degrees. The phase lag Φ which is

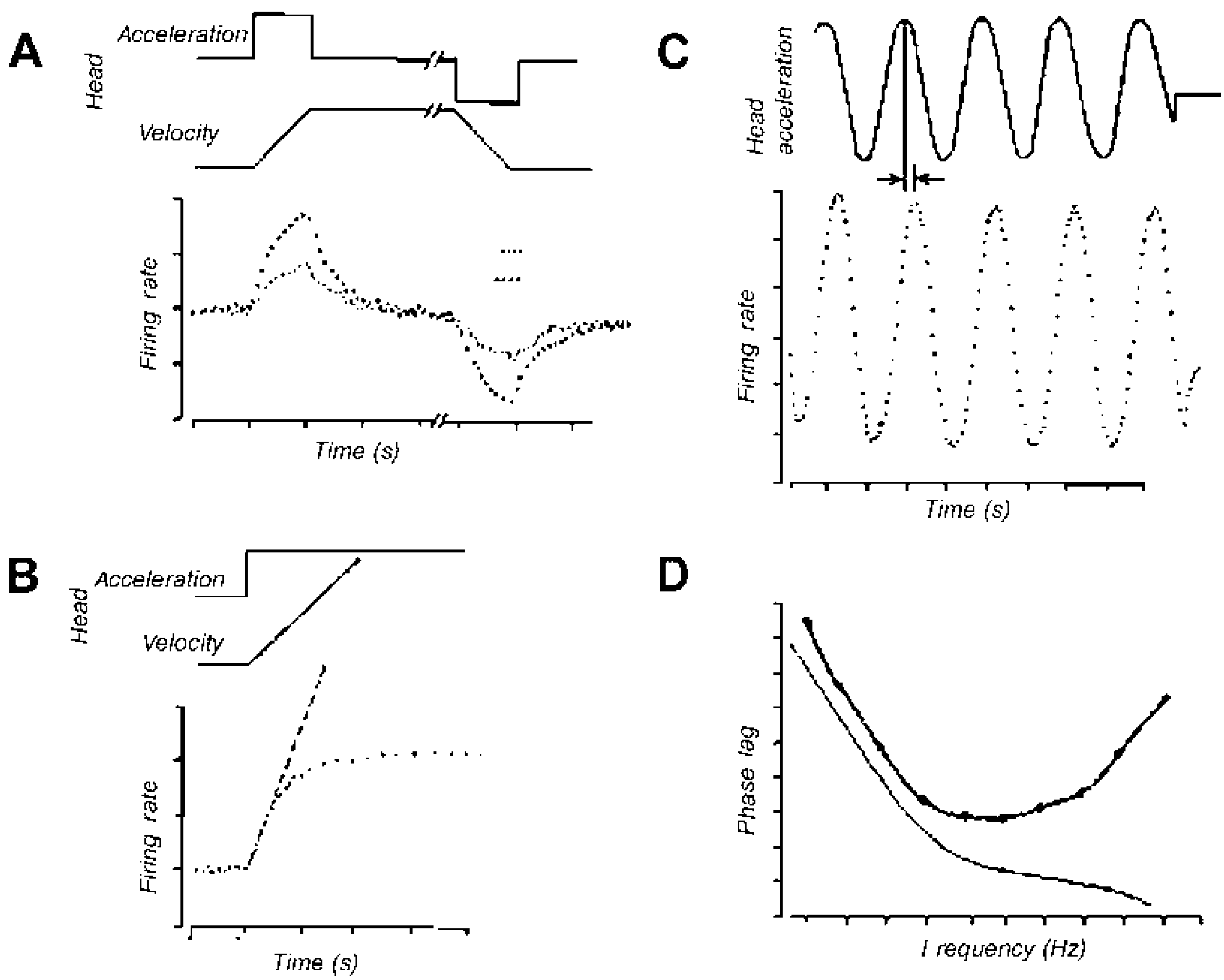


Figure 33–1 *(A–D) Relationship between head acceleration, head velocity, and vestibular neuron firing. Individual graphs are discussed in the text.*

shown in panel C of **Fig. 33–1**, actually is smaller than 90 degrees, so the integration is incomplete. Panel D of **Fig. 33–1** summarizes the behavior of many canal afferents seen in many species, including primates. Some fibers do show a 90-degree phase lag over the frequencies of head movements that we commonly encounter, but most have a smaller phase lag. Indeed, some can more appropriately be said to encode the input acceleration. As we learn more about the other vestibular reflexes (vestibulocolic, vestibulospinal), we may find the role that these acceleration signals play in the sense of balance.

THE BRAINSTEM AND BEYOND: THE VESTIBULO-OCULAR REFLEX

We now consider the means by which the vestibular signals reach the target areas of the brainstem that control eye movements. Primary afferents project almost exclusively to the ipsilateral four vestibular nuclei on the floor of the fourth ventricle at the pontomedullary junction. These are the superior, medial, lateral, and descending vestibular nuclei. As a rough rule, canal inputs project to the more rostral parts of the superior, medial, and lateral nuclei. The maculae otolith inputs project most heavily to the caudal medial and descending nuclei.

We know that stimulation of the right horizontal canal (as by head turning to the right) must result in compensatory conjugate eye movements to the left. This requires activation of the left lateral rectus and right medial rectus. We can trace here a three-neuron reflex arc going to the left lateral rectus and a three-neuron reflex arc to the right medial rectus. Simultaneously, the right lateral and the left medial recti must be inhibited. These pathways can be traced as well. Thus each canal gives rise to two excitatory and two inhibitory pathways, for a total of 12 pathways. Note that eight of these travel in the medial longitudinal fasciculus. Lesions here, like the classic occurrence of a multiple sclerosis plaque causing internuclear opthalmoplegia, will disrupt the machinery that produces the normal conjugate eye movements of the VOR.

We have considered the "hardware," or the wiring of the system. But what does the "software" do? That is, what sort of signal processing occurs in these secondary and higher order neurons that is of clinical importance? Recall that we found that afferents carried an incompletely integrated signal. If one supposed that eye movements needed to be generated from these signals, then it becomes clear that one of the tasks for the brainstem is to complete the integration. Thus the brainstem acts as a velocity to position integrator. There are more circuitous pathways for vestibular input to reach the ocular motor centers than the simple three-neuron arcs we have seen. Some of these go through the reticular formation and the nucleus prepositus hypoglossi, where the integrator function may lie. However, this integrator function is not unique to vestibular reflexes, and it is shared with other eye movement systems. For example, when we move and hold our gaze to one side, the integrator takes a pulse signal to move the eye and integrates it to get a prolonged signal that holds the eye in the new position.

One of the consequences of integrator dysfunction is the inability to hold eccentric gaze. The failing or "leaky" integrator allows the eye to drift slowly back to the neutral position from wherever it is in the orbit. This explains Alexander's law, which states that nystagmus (vertigo) will be augmented when looking in the direction of the fast phase. When looking toward the fast phase, the slow drift in the opposite direction adds to the slow phase, augmenting the nystagmus. When looking toward the slow phase, however, the drift is in the opposite direction and subtracts from the slow-phase velocity.

Another failing of the afferent input is that it dies away too quickly. The time constant of the cupula, a measure of how fast it returns to its resting position after a displacement, is ~13 seconds. A VOR based on this would do fine during high-frequency head movements, but at lower frequencies it would cause the eye to complete its motion too quickly, giving it a large phase lead at low frequencies. We will see this when we discuss rotary chair testing.

The brainstem makes up for this deficit by prolonging the vestibular input signal, an operation known as velocity storage. This may be accomplished by recurrent loops in the VOR circuitry that add up the signal and "store" it over time. Note: This is not the same as the velocity to position integrator we discussed previously, a common misconception, but like the integrator, velocity storage is shared with other eye movements. So, for example, optokinetic nystagmus dies away slowly as an "afternystagmus" because of the effect of velocity storage. Velocity storage is commonly crippled on the side of a unilateral labyrinthine hypofunction. Thus a prolonged sinusoidal input will be stored on the intact side, but not on the lesioned side. This is the origin of post-headshaking nystagmus.

THE VESTIBULOCOLIC AND VESTIBULOSPINAL REFLEXES

The vestibular system also keeps us upright with respect to gravity. Linear acceleration signals from the otolith organs are processed in the medial and lateral vestibular nuclei, which give rise to the medial and lateral vestibulospinal tracts (MVST and LVST, respectively). The MVST provides input to the neck muscles, and this explains why vestibular lesions can cause a head tilt. The LVST descends to the thoracic and lumbar spinal cord to drive the extensor (antigravity) muscles of the lower limbs.

THE CEREBELLUM AND THE VESTIBULAR SYSTEM

We have seen how the vestibular system acts like a machine to drive eye movements in the canal planes. There may be circumstances when we wish to turn off the VOR, as, for example, when we are following a moving target by moving our head. If the VOR stayed on in this circumstance, the eyes would be driven off in the wrong direction. Control of the VOR for these situations is provided by the cerebellum, in particular by the flocculus.

The canal input is sent not only to the interneuron in the reflex arc but also to the cerebellum. The Purkinje's cells of the cerebellum provide a copy of the signal, but now as an inhibitory signal, to the same interneuron. The result is a cancellation of the VOR. Probably the most common situation in which this occurs is when we move our head to shift our gaze. Here we do not want the eye to remain stationary in space, but to move along with the head. (In actuality, the eye gets on target before the head, and the VOR is turned back on at the end of the head movement to keep the eye there.)

The cerebellum is critical to vestibular adaptation. When there is a loss of vestibular function, leading to a decreased input from one side, the cerebellum adjusts the central circuits to cancel the imbalance. The cerebellum also adjusts the system to compensate for changes in our vision (e.g., presbyopia, new eyeglasses), so that the VOR does not cause the image to overshoot or undershoot the fovea because of the magnification change.

Suggested Readings

Baloh RW and Halmagyi GM. Disorders of the Vestibular System. New York, Oxford University Press; 1996

Carey JP and Della Santina CC. Principles of applied vestibular physiology. In Cummings C.W., Otolaryngology-Head and Neck Surgery, Fourth Edition, 2005. Philadelphia, Elsevier Mosby 2005; 3115–3159

Ewald JR. Physiologische Untersuchungen uber das Endorgan des Nervus Octavus. In Wiesbaden, Germany. Bergmann; 1892

Highstein SM, Fay RR, Popper AN, eds. The Vestibular System (Springer Handbook of Auditory Research, Vol 19). New York, Springer-Verlag; 1995

Steinhausen W. Über die Beobachtung der Cupula in den Bogengängsampullen des Labyrinths des lebenden Hechts. Pflügers Arch Ges Physiol 1933; 232, 500–512

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Which vestibular organ has its kinocilia located closest to the utricle?
   - A. Saccule
   - B. Posterior canal
   - C. Superior canal
   - D. Horizontal canal
2. The vestibular system in one labyrinth works
   - A. As an isolated sensor of movement
   - B. In conjuction with the matched sensory organs in the opposite labyrinth
   - C. Primarily to sense velocity
   - D. As a high-frequency movement sensor
3. Ewald's first law postulates that
   - A. Nystagmus is generated in the plane of the semicircular canal being stimulated.
   - B. Nystagmus is strongest when gazing in the direction of the fast phase of nystagmus.
   - C. Only vertical nystagmus is nontorsional.
   - D. Unilateral weakness may be detected by the presence of nystagmus.

Chapter 34

TESTING BALANCE AND THE VESTIBULAR SYSTEM

Hinrich Staecker

Testing balance is a complex subject due to the variety of sensory systems involved in the perception of balance. Testing falls into two main groups: tests that rely on the activation of the vestibulo-ocular reflex (VOR; i.e., electronystagmogram and rotation testing) and tests of global balance (posturography). The electronystagmogram (ENG) is actually a test battery that screens for central as well as peripheral vestibular dysfunction. Finally, there are a variety of experimental vestibular test procedures that test otolith function and cervical vertigo and attempt to test directly activation of the central vestibular system. The basic principles that underlie vestibular testing are the evaluation of the VOR (see Chapter 33), through the assessment of the output response induced by stimulation of the horizontal semicircular canal. The basics of the VOR response are shown in **Fig. 34–1**.

THE ELECTRONYSTAGMOGRAM

The fundamental test of balance is the ENG. The ENG test battery relies on the recording of nystagmus (vertigo) after a variety of stimuli. This is most frequently accomplished by placing electrodes that measure the shift in the corneoretinal potential. The cornea has a positive (+) charge, and the retina has a negative (−) charge. The movement of the eye creates a movement of the dipole that can be recorded and amplified. Calibration is accomplished by comparing the electrode output to eye movement induced by looking at a target point. By knowing the angle of eye movement divergence and recording the change in the eye dipole on a fixed speed chart recorder, the eye movement speed can be calculated (**Fig. 34–2**). By convention, an upward deflection on the chart recorder represents a rightward deviation, and a downward - deviation represents a leftward eye deflection. Standard electrode placement also allows placement of a vertical electrode that records upward or downward eye movement. ENG electrodes do not record torsional eye movements because the dipole does not shift. This makes it difficult to record nystagmus that has a torsional component, such as benign paroxysmal positional vertigo (BPPV). Recording of eye movement with infrared video cameras allows the recording of nystagmus in all directions. Assessment of vestibular defects using ENG consists of three phases. Initially, the presence of spontaneous nystagmus with eyes open and closed is tested, looking for congenital nystagmus as well as suppressible and nonsuppressible nystagmus. Testing of pursuit, saccade testing, and optokinetic testing are then done to check for central dysfunction (**Table 34–1**). Positional and positioning nystagmus are then evaluated. A review of expected eye movements

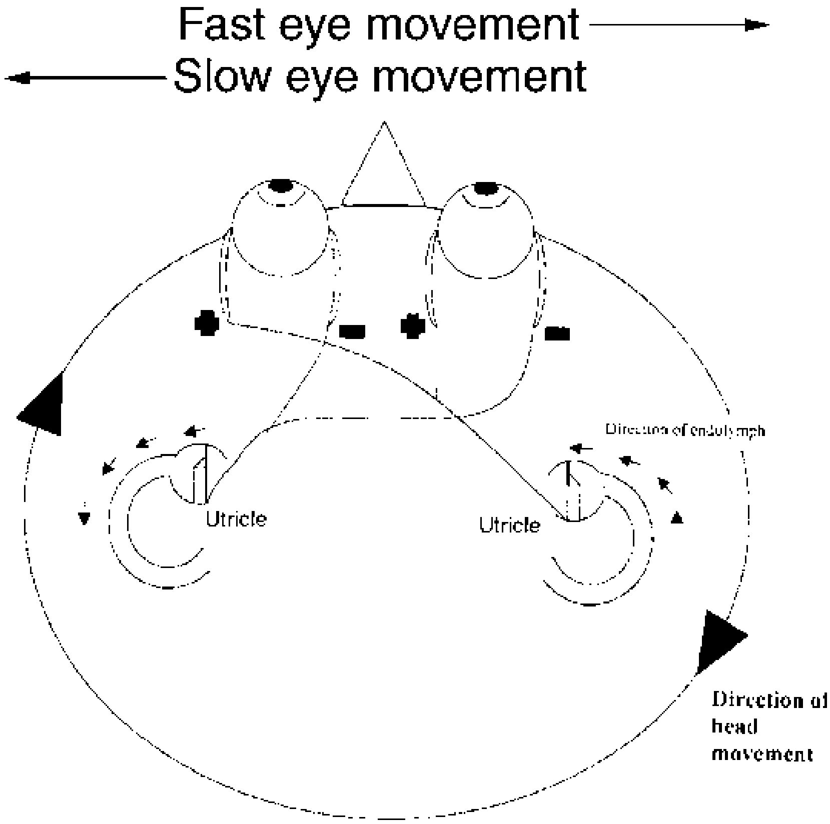


Figure 34–1 *A schematic representation of eye movement in response to a rotational stimulus. Seen from above, a clockwise or rightward head movement results in deflection of the horizontal canal kinocilium toward the utricle on the right and away from the utricle on the left. The right hair cells are excited and increase their firing rate, whereas the left hair cells are inhibited and decrease their firing rate. The firing rate of the right vestibular nucleus increases, whereas the firing rate of the left vestibular nucleus decreases. This leads to signals to the oculomotor nuclei of the III and VI nerves (discussed in greater detail in Chapters 27 and 33). Essentially, the left lateral rectus and the right medial rectus are stimulated to contract, whereas the left medial and right lateral rectus muscles are allowed to relax. This results in a slow eye movement to the left and a fast saccadic movement to the right. Because nystagmus is defined by the fast phase, this results in a right-bearing nystagmus. Warm water irrigation on the right side would have a similar effect.*

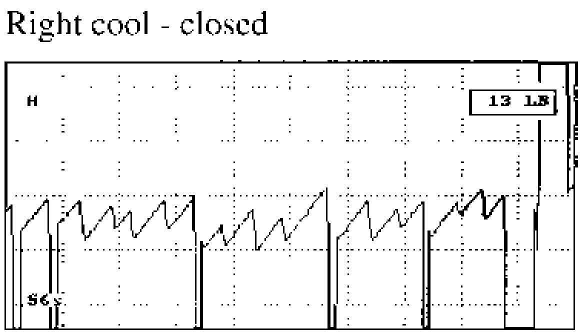
induced by stimulation of the various canals is given in **Table 34–2**.

THE CALORIC TEST

The caloric test evaluates the function of the horizontal semicircular canals. A variety of stimuli (open loop, closed loop, air) are used to change the temperature of the external auditory canal (EAC). Open-loop water calorics are considered the most reliable, with the lowest test–retest variability. The test is performed with the patient lying at a 30-degree angle, which places the horizontal semicircular canal perpendicular to the ground. Because the horizontal canal is closest to the EAC, the change in EAC temperature during irrigation with water is transmitted to the horizontal canal. The change in temperature induces a change in density of the endolymph, resulting in flow of endolymph and displacement of the cupula. The stimulus provided by calorics is approximately equivalent to a rotational stimulus of 0.025 Hz. Normal walking and head turns yield stimuli in the 4 to 8 Hz range. Alternating warm (44°C) and cool (30°C) stimuli result in endolymph flow in opposite directions, yielding nystagmus in opposite directions. A cold irrigation causes an ampullofugal flow, deflecting the kinocilium away from the utricle and inhibiting the spontaneous firing rate of the hair cells. A warm irrigation causes ampullopetal flow, thereby deflecting the kinocilium toward the utricle and exiting and increasing the firing rate of the hair cells. This increase in the rate of firing results in stimulation of the ipsilateral III nerve and contralateral VI nerve nucleus. This induces a deviation of the eyes toward the opposite ear (slow phase of nystagmus) and a saccadic correctional movement of the eyes toward the stimulated ear (fast phase of nystagmus). By convention, the direction of nystagmus is defined by the fast phase. Therefore, a warm irrigation in the right ear will produce a right-bearing nystagmus, which is shown on the chart recorder as an upward deflection. The mnemonic COWS describes the direction of nystagmus response after a caloric stimulus: *c*old water irrigation nystagmus bears toward the *o*pposite ear; *w*arm water irrigation nystagmus bears toward the *s*ame side ear. The normal rate of nystagmus after caloric stimulation ranges from 8 to 80 degrees per second. When both ears are alternately irrigated with warm and cold water, nystagmus rates can be determined, and the function of one horizontal canal with relationship to the other can be determined. The Jongkees formula

$$[(L30 + L44) - (R30 + R44)]/(L30 + L44 + R30 + R44),$$

where L is left, R is right, 30 is 30°C irrigation, and 44 is 44°C, describes the presence of unilateral weakness;



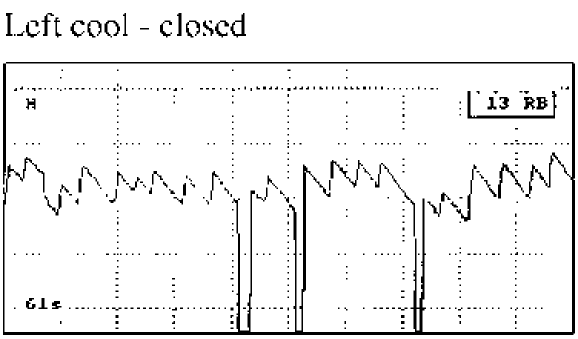


Figure 34–2 *A basic electronystagmography tracing shows one channel of a recording. Upward deflections represent rightward eye movement, and downward deflections represent leftward eye movement. The more perpendicular deflections represent fast eye movement, whereas the shallow sloped deflections represent the slow component of eye movement. The equipment is calibrated such that 1 degree of eye displacement equals 1 mm of pen displacement. Given a known paper speed (usually 10 mm per second), one can determine the slow-phase velocity of the nystagmus.*

that is, there is less nystagmus produced by one ear, which represents a peripheral vestibular weakness. A second formula

$$[(L30 + R44) - (R30 + L44)]/(L30 + L44 + R30 + R44)$$

describes directional preponderance (i.e., the nystagmus tends to bear more to the right or left). This is a nonlocalizing finding. In the absence of a caloric response, an ice water caloric irrigation can determine if there is any residual function in the labyrinth.

ROTATIONAL TESTING

An alternate way of testing the VOR is to stimulate the horizontal canal by rotation. By rotating a patient in a motorized chair and measuring the compensatory eye movements over multiple rotations at different frequencies, a graph of vestibular function (horizontal canal) at more physiological stimulus frequencies can be obtained. This test can be performed in several different ways, but most commonly it is done with a clockwise and counterclockwise rotational stimulus. The stimulus can be described as a sine wave with a known frequency and angle of displacement. The time it takes to complete one cycle of clockwise and counterclockwise rotation is the period (T). The stimulus frequency is the number of oscillations of the chair per second and is measured in Hertz (i.e., a 1 Hz stimulus is a 1 cycle per second stimulus). As the frequency increases, the time it takes to complete a cycle decreases; thus frequency is inversely related to period. For a given period, changing the maximum angle of chair displacement will change the chair acceleration and velocity; thus displacement is usually kept constant. Frequencies tested range from 0.01 to

TABLE 34–1 SUMMARY OF ENG TESTING AND CLINICAL SIGNIFICANCE

ENG test	Abnormality	Localization
Saccade	Dysmetria	Cerebellum
	Slowing	Central
Tracking	Saccadic	Central
	Disorganized	
Optokinetic	Asymmetry	Central
Positional	Nystagmus (eyes open, fixed direction)	Usually central
	Nystagmus (eyes open, changing direction)	Central
	Nystagmus (eyes closed, fixed direction)	Peripheral
	Nystagmus (eyes closed, changing direction)	Central
Hallpike	Rotatory, upbearing	BPPV, posterior canal
	Rotatory, downbearing -onset after latency, fatigable	BPPV, superior canal
Supine-head turn lateral	Horizontal toward downward ear	BPPV, horizontal canal
Calorics	Unilateral weakness	Peripheral lesion
	Bilateral weakness	Peripheral lesion/poor Irrigation
	Directional preponderance	Peripheral or central

BPPV, benign paroxysmal positional vertigo; ENG, electronystagmography

TABLE 34–2 STIMULUS PATTERNS OF THE VESTIBULAR END-ORGANS

Horizontal canal	Kinocilium turned toward utricle-utriculopetal flow is excitatory
Superior canal	Kinocilium turned away from utricle-utriculopetal flow is inhibitory
Posterior canal	Kinocilium turned away from utricle-utriculopetal flow is inhibitory
Horizontal canal stimulation	Stimulates ipsilateral medial rectus and contralateral lateral rectus
Horizontal canal inhibition	Releases contralateral medial rectus and ipsilateral lateral rectus
Superior canal stimulation	Stimulates ipsilateral superior rectus and contralateral inferior oblique
Superior canal inhibition	Releases ipsilateral inferior rectus and contralateral superior oblique
Posterior canal stimulation	Stimulates ipsilateral superior oblique and contralateral inferior rectus
Posterior canal inhibition	Releases ipsilateral inferior oblique and contralateral superior rectus

1 Hz and include those frequencies that are harmonics (multiples) of the fundamental test frequency. This tests response over the low- to midfrequency range within the linear range of the system's dynamic range. For each frequency point, testing must be performed over an extended time period to allow the eye movement response to reach a steady state. Alternate testing protocols include sum of sines, which combines multiple frequencies, pseudorandom, and velocity trapezoids, all of which have distinct advantages and disadvantages. Testing is performed with eyes open and in the dark, which yields stronger and more reliable responses. Subject attention is important for collecting accurate data.

The nystagmus response is recorded either with ENG electrodes or with an infrared video camera averaged and transformed into a sinusoid and can be described in terms of position, velocity, and acceleration. As shown in **Fig. 34–1**, the slow eye movement should be in the opposite direction of chair movement. The speed of the slow component of nystagmus can then be measured in several different ways. Phase describes the temporal difference between the stimulus waveform (chair movement) and the output waveform (eye movement), and gain describes the ratio of stimulus to eye movement output. At high frequencies the eye velocity matches the chair velocity (therefore, there is little difference between timing of the stimulus and response), the phase is low, and gain approaches one. At lower frequencies this relationship breaks down. (This is demonstrated in the plot of a normal rotation test seen in **Fig. 34–3**.) To understand why this occurs, we must briefly review the VOR in greater detail. The processing of the VOR can be broken down into two segments. Initially, angular acceleration is mechanically detected by the deflection of the cupula. Mathematical models have shown that deflection of the cupula is proportional to the velocity of head movement at frequencies greater than 0.1 Hz. This relationship breaks down at lower frequencies, requiring the addition of central processing to allow accurate eye movements. The stimulation of each vestibular end-organ results in a change in activity of the vestibular nucleus; this yields stimulation of agonist muscles and inhibition of antagonist muscles produced by each side. Several asymmetries are introduced to the system that

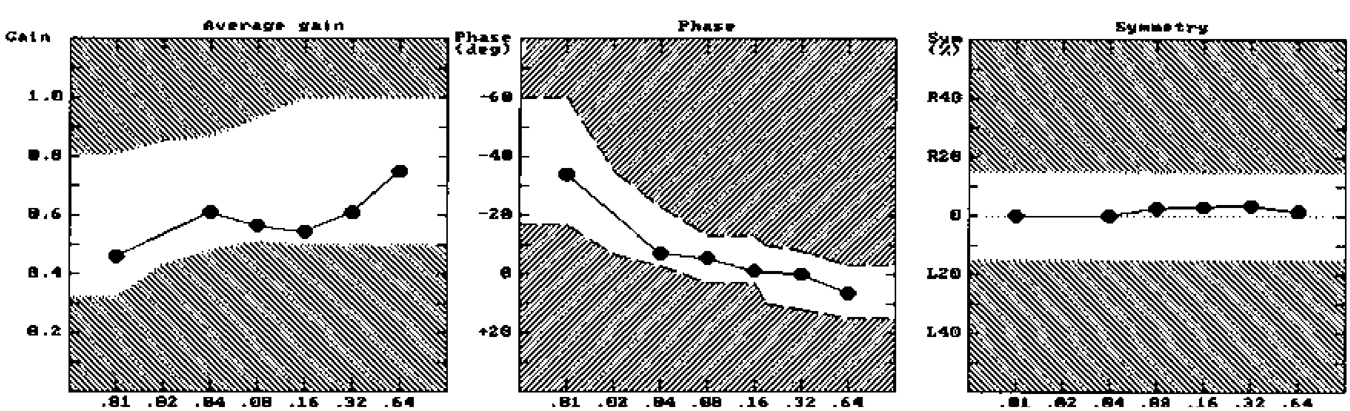


Figure 34–3 The results of a rotation test using sinusoidal harmonic stimulation at a frequency range of 0.01 to 0.64 Hz. The results are shown within 2 and 2.5 standard deviations. Both gain and phase have a linear pattern between 0.1 and 1 Hz. At lower frequencies the phase lead increases. Gain represents the relationship between maximum stimulus and maximum eye velocity, whereas phase represents the timing difference between the stimulus and the output wave (see text). Symmetry of eye movement (predominance of rightward and leftward nystagmus) is shown in the rightmost graph.

TABLE 34–3 SUMMARY OF ROTATION TEST FINDINGS

Gain	↑	↓	↓/Nl	↓/Nl	↓
Phase	Nl	↑	Nl	↑	↑
Diagnosis	Central/poor tasking	Bilateral Hypofunction	Poor tasking	Unilateral compensated	Unilateral peripheral

must be corrected by central processing. Initial asymmetries are introduced by the mechanical properties of the cupula, as already discussed. Next, there is an asymmetry of rate sensitivity between the left and right side. Again, looking at the horizontal semicircular canal, each hair cell on each side has a baseline firing rate. A positive stimulus will enhance the firing rate of one side while decreasing the firing rate of the other side. The increase of firing is ultimately limited by the refractory period of the vestibular neurons. However, inhibition is possible only to a lesser degree because a hair cell cannot decrease its spontaneous discharge rate below zero. The asymmetry is overcome by having the system function in a push-pull fashion and by introducing significant central integration of bilateral signals. This makes the phase measurement particularly sensitive to unilateral abnormalities of the peripheral vestibular system. Partial or complete loss of information from one side creates progressively enlarged phase abnormalities that are not completely corrected by central compensation.

Comparison of peak eye velocity to peak chair velocity yields the gain. Gain is initially decreased in a peripheral vestibular lesion. Studies have shown that spontaneous firing rates of the vestibular nucleus increase after labyrinthectomy, thus allowing gain to return to normal after a peripheral vestibular lesion. Thus test results showing a decrease in gain and an increase in phase imply an acute or uncompensated lesion, whereas test results showing normal gain and increased phase imply that compensation has taken place. Symmetry can be determined by comparing the leftward and rightward eye movement velocities by the following formula:

$$((Vr - Vl)/(Vr + Vl) \times 100)$$

A graphical representation of gain, phase, and symmetry is then produced over a range of frequencies (**Fig. 34–3**). As can be seen in **Fig. 34–3**, phase retains a linear relationship over mid- to high-range frequencies and increases at low frequencies. This is due to the mechanical properties of the cupula and the effect of central processing. The frequency at which the phase loses its linear relationship and increases is called the corner frequency and is the frequency at which the phase lead is equal to 45 degrees. The time constant (τ) is a measure of the relationship between the stimulus, deflection of the cupula, and the effect of central processing. The value of τ may be determined by taking the inverse of the corner frequency times 2Π ($\tau = 1/2\Pi f$). The time constant is a sensitive indicator of vestibular dysfunction because, as already discussed, aberrations of peripheral input affect the symmetry of central processing, thereby altering the phase and time constant.

Compared with the caloric test, which essentially checks only one frequency, the rotation test looks at a range of values, just like an audiogram tests a range of hearing frequencies. Interpretation of test results requires normative data for various age groups. Overall, using a battery of tests that include both caloric testing and rotation yields the greatest diagnostic sensitivity. Several different patterns of findings and their interpretations are reviewed in **Table 34–3**.

VISUAL-VESTIBULAR INTERACTION

Combining rotation testing with visual stimuli generates a sensitive test of central balance disorders. The test subject is exposed first to standard rotation testing to determine the VOR gain, phase, and symmetry. The fixation test considers the ability to suppress nystagmus by focusing on a fixed target. This allows one to determine the fixation index, which is a sensitive indicator of central nervous system dysfunction. Optokinetic nystagmus is tested by exposing the subject to a rotating pattern of vertical stripes that surround the patient's entire visual field. The visual VOR (VVOR) is tested by performing harmonic sinusoidal rotation of the patient with eyes open and a stationary optokinetic drum.

POSTUROGRAPHY

Because balance is a combination of vestibular, visual, and proprioceptive sensations, several forms of posturography have been designed to evaluate global balance function. The most commonly used today is computed dynamic posturography. Patients with dysequilibrium of unknown etiology, a history of recurrent falls, a history of head trauma, or persistent dizziness despite a negative workup, as well as suspected malingerers, are candidates for this test. The test subject stands on a "force plate" that is able to detect motion and weight distribution. A computer program then changes the position of the floor or the visual horizon. Using sets of normative data, the patient's response to adverse balance situations can be determined. Depending on the combination of stimulus conditions, testing can suggest whether a balance disorder is due to proprioceptive, visual, vestibular, or a combination of problems. This portion of the test is known as the sensory organization test (SOT). Overall, six conditions are tested, combining platform fixed or moving with eyes open or closed and fixed or moving platform with moving visual surround. Conditions 5 and 6 (eyes closed, platform moving, and platform and visual surround moving) test for vestibular dysfunction. Posturography does not provide diagnostic or localizing (e.g., peripheral vestibular vs cerebellar) information, only an assessment of balance function. Additional tests are used to evaluate the patient's center of gravity and whether balance is maintained using movements of the ankle or hip. The CDP setup is able to sense the speed of response to a forward or backward translation or tilt, allowing a calculation of reflex speed. Again, using normative data, abnormal reflex times can be identified, suggesting neuropathies of central motor disorders. Falls and failure of various components of the test occur in repeatable patterns. This has allowed the testing to be used to identify functional patients.

TESTS UNDER DEVELOPMENT

Traditional VOR testing is limited to examination of the horizontal semicircular canal. The maculae of the utricle and saccule contribute significantly toward balance function, acting as accelerometers and gravity sensors. Saccular function is currently being investigated using a sound stimulus that triggers contraction of the sternocleidomastoid muscle. Other tests have focused on factors such as perception of the visual vertical and measuring the influence of otolith function on the VOR.

Suggested Readings

Baloh R, Halmagyi M, eds. Disorders of the Vestibular System. New York: Oxford University Press; 1996

Baloh R, Honrubia V. Clinical Neurophysiology of the Vestibular System. Philadelphia: FA Davis; 1990

Barber H, Stockwell C. Manual of Electronystagmography. St. Louis: CV Mosby; 1976

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. Normal electronystagmography electrodes are unable to detect
   - A. Direction-changing nystagmus
   - B. Vertical nystagmus
   - C. Horizontal nystagmus
   - D. Rotatory nystagmus
2. Posturography allows localization of vestibular disease to
   - A. The cerebellum
   - B. The otolith organs
   - C. The visual system
   - D. None of the above
3. Which of the following tests has the highest test–retest reliability?
   - A. Caloric irrigation with air
   - B. Caloric irrigation with a closed loop system
   - C. Caloric irrigation with an open loop system
   - D. All of the above

Chapter 35

MORPHOPHYSIOLOGY OF THE FACIAL NERVE

K. Paul Boyev and Adrien A. Eshraghi

Knowledge of the facial nerve—its course, function, and rehabilitation—influences decision making in multiple clinical subspecialties, including otology, facial plastic surgery, and head and neck oncology. Therefore, it is important to understand, the complex anatomy and physiology of this nerve in health and disease.

By convention, the facial nerve is divided into intracranial, intracanalicular, labyrinthine, geniculate, tympanic (horizontal), mastoid (vertical), and extracranial segments. The main trunk of the extracranial segment is then subdivided first at the pes anserinus and then further as it divides into the frontal, zygomatic, buccal, marginal mandibular, and cervical branches.

CENTRAL ANATOMY

The facial nerve has sensory, motor, and autonomic components. The 10,000 axons of the facial nerve roughly correspond to the 7000 cell bodies in the facial motor nuclei and 3,000 fibers arising from efferents in the salivatory and lacrimal nuclei, and afferents to the nucleus tractus solitarius. Corticobulbar fibers originating in somatomotor cortex cell bodies project through the internal capsule to caudal pons, where they decussate before synapsing with both ipsilateral and contralateral facial nerve motor nuclei. The dorsal portion of the facial motor nucleus receives both ipsilateral and contralateral innervation and supplies the superior facial mimetic musculature; the ventral facial motor nucleus receives

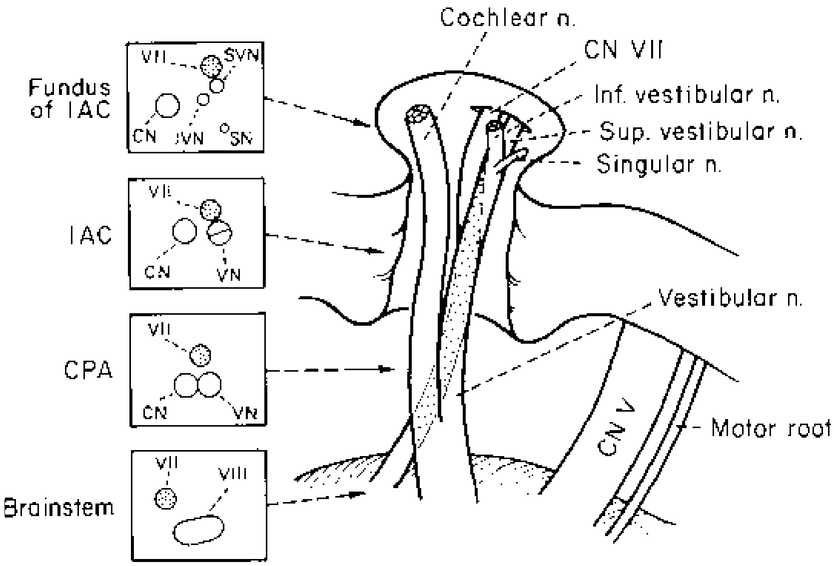


Figure 35–1 *Rotation of the facial nerve in relationship to the eighth nerve in the cerebellopontine angle (CPA). Cochlear nerve (CN); fifth cranial nerve (CN V); seventh cranial nerve (CN VII and VII); eighth cranial nerve (VIII); internal auditory canal (IAC); inferior vestibular nerve (IVN); singular nerve (SN); superior vestibular nerve (SVN); vestibular nerve (VN). (From May M, Schaitkin BM. The Facial Nerve, May's 2nd ed. New York: Thieme Medical Publishers; 2000:33, Fig. 2–12. Reprinted with permission.)*

only contralateral fibers, innervating the lower portion of the face: this is the basis for the sparing of forehead function seen with a unilateral upper motor neuron lesion. The superior salivatory and lacrimal nuclei are also located in the caudal pons, and it is from these nuclei that parasympathetic general visceral efferent nerve fibers originate. These fibers supply the lacrimal gland and extraparotid salivary glands.

One finds the termini of afferent fibers in the nucleus tractus solitarius (NTS) and the spinal trigeminal tract. The chorda tympani nerve is composed of special visceral afferent fibers whose cell bodies are located in the geniculate ganglion and which project to the NTS. Sensation from the skin of the external auditory canal and postauricular area is conveyed via general sensory afferents terminating in the spinal trigeminal tract.

Cerebellopontine Angle

The facial nerve traverses the cerebellopontine angle in close association with the cochleovestibular nerve and the intermediate nerve of Wrisberg, which carries parasympathetic, general visceral efferents to the salivary and lacrimal glands, as well as special visceral afferent fibers from the chorda tympani and general sensory afferents from the external auditory canal and postauricular area. This segment of the nerve derives its vascular supply from the labyrinthine artery off the anterior inferior cerebellar artery. As the facial and cochleovestibular nerves enter the internal auditory canal, the vestibular and cochlear nerves rotate 90 degrees such that the cochlear nerve, which leaves the pons posterior to the facial nerve, ultimately assumes an inferior and slightly anterior relationship to the facial nerve as it progresses laterally (**Fig. 35–1**).

Internal Auditory Canal

Fibers of the nerve of Wrisberg, frequently called the nervus intermedius when visible in the internal auditory canal (IAC), fuse to form a single bundle with the motor fibers as they move laterally. Near the lateral extreme of the IAC, the vertical crest (Bill's bar) marks an anterior and posterior division of this canal; the transverse crest, located anteriorly, divides the canal into superior and inferior divisions. Superior to the transverse crest one finds the facial nerve anteriorly, and the superior vestibular nerve is separated posteriorly by Bill's bar. Inferiorly, the transverse crest marks a separation of the facial nerve from the cochlear nerve, which are both anterior structures (see Chapter 22). The facial nerve's narrowest portion occurs as it exits the IAC at the meatal foramen, which is on average no wider than 0.68 mm.

Intratemporal Course of the Facial Nerve

Within the temporal bone, the facial nerve courses through the fallopian canal, measuring 30 mm in length between the meatal and stylomastoid foramina. The labyrinthine segment of the facial nerve originates distal to the meatal foramen and is the shortest segment, 3 to 4 mm in length, before reaching the geniculate ganglion (the so-called first genu). At the geniculate ganglion, there is a dehiscence in the floor of the middle cranial fossa called the facial hiatus, through which the greater superficial petrosal nerve (GSPN) emerges (**Fig. 35–2**).

Cell bodies for the chorda tympani and for the sensory nerves of the GSPN are found in the geniculate ganglion. In addition, preganglionic secretory fibers pass through the geniculate ganglion on their way to the sphenopalatine ganglion via the GSPN.

The tympanic segment spans the 12 to 13 mm between the first and second genu of the facial nerve. Making a sharp posterior and lateral bend at the geniculate ganglion, it maintains a superior relationship to the canal of the tensor tympani muscle and cochleariform process. In 30 to 50% of temporal bone specimens, the segment in the medial wall of the middle ear cavity can be dehiscent. Continuing, the nerve passes medial to the incus and inferior to the horizontal semicircular canal, where it gives off a branch that innervates the stapedius muscle. Posterior to the oval window, the tympanic segment of the facial nerve enters the sinus tympani and turns inferiorly at the second genu, forming the mastoid segment of this nerve.

The facial nerve during its descent through the mastoid gives off two more branches: the sensory auricular branch supplying the external ear and the chorda tympani nerve (**Fig. 35–3**). The mastoid segment continues vertically for a distance of 12 to 20 mm where the facial

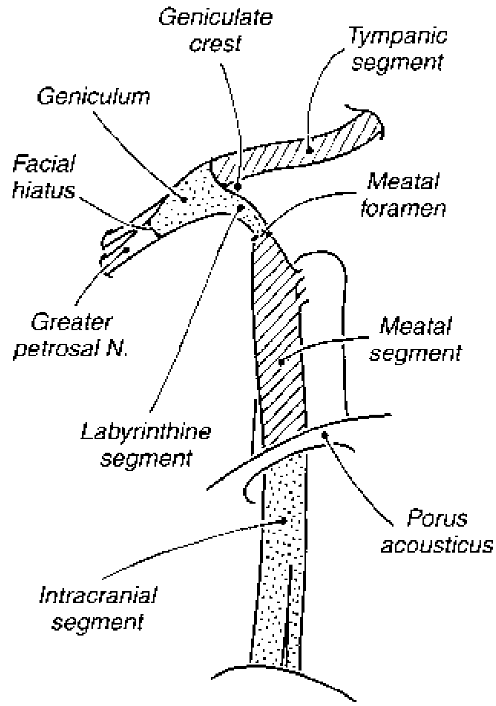


Figure 35–2 *Intratemporal course of the facial nerve. (From Fisch H et al. Microsurgery of the Skull Base. New York: Thieme Medical Publishers; 1988. Reprinted with permission.)*

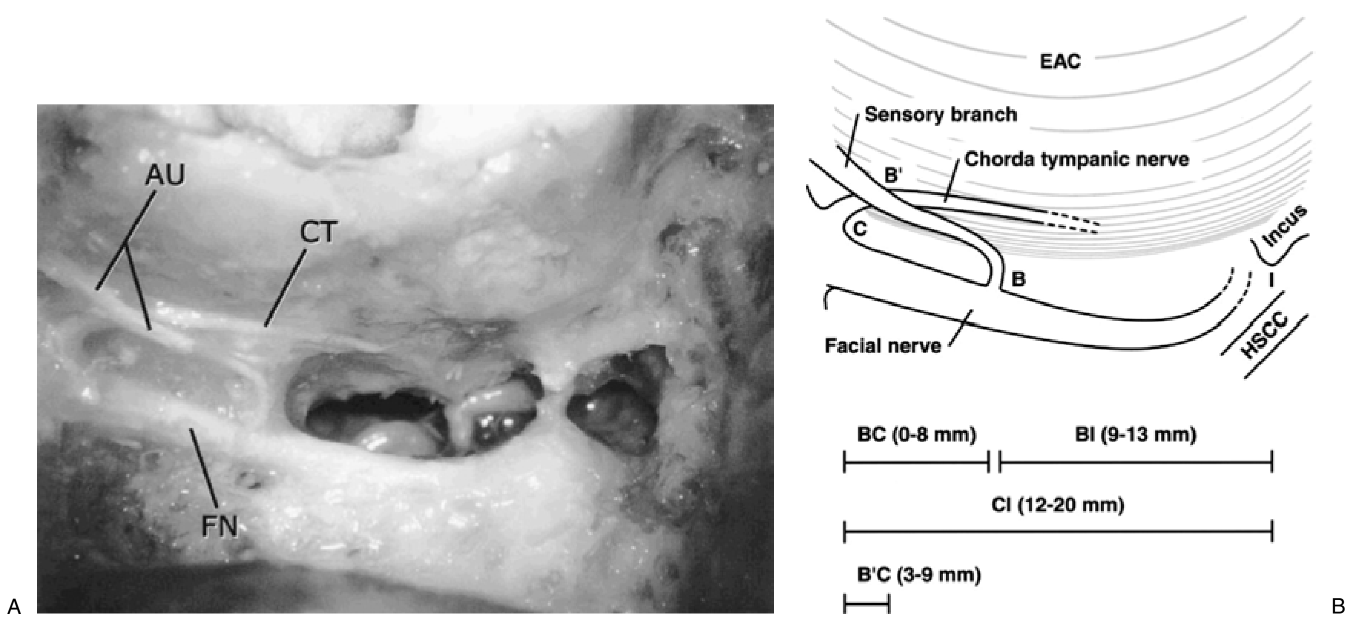


Figure 35–3 *Surgical anatomy of the mastoid portion of the facial nerve.* ***(A)*** *Macroscopic photograph of a left human temporal bone after dissection of the mastoid cavity and the facial recess with identification of the facial nerve (FN) and the chorda tympani (CT) and auricular (AU) branches of this cranial nerve.* ***(B)*** *Diagram of the course of the facial nerve through the mastoid region of the temporal bone, with identification of the tip of the short process of the incus (I); the chorda tympani nerve, origin at the fallopian canal (C); the origin of the auricular branch of the facial nerve (B); and the point of intersection between the chorda tympani and the auricular branch of the facial nerve (B'). HSCC, horizontal semicircular canal. (From Eshraghi AA et al. Sensory auricular branch of the facial nerve. Otol Neurotol 2002;23: 393–396. Reprinted with permission.)*

nerve exits at the stylomastoid foramen. In children younger than 2 years, the facial nerve can be dangerously superficial at its exit because it is not yet protected from lateral damage by either the tympanic ring or the mastoid tip. With growth of the temporal bone, the stylomastoid foramen comes to lie between the mastoid tip laterally and the styloid process medially.

Extracranial Course of the Facial Nerve

As the facial nerve exits the stylomastoid foramen, it gives nerve branches to the posterior belly of the digastic, stylohyoid, and postauricular muscles. The main trunk of the facial nerve immediately plunges into the parenchyma of the parotid gland, dividing it into superficial and deep lobes. It courses for approximately 2 cm before bifurcating at the pes anserinus. Distal to the bifurcation, the frontal, zygomatic, and buccal branches interconnect in a highly variable plexus composed of thin nerve filaments. The marginal mandibular and cervical branches of the facial nerve are less likely to form such interconnections (**Fig. 35–4**).

Motor End Plate

Facial nerve fibers interface with striated mimetic musculature at the motor end plate in an acetylcholine-dependent synapse. This neuromuscular junction resembles a synapse at the nerve terminal, but the distal portion is a specialized muscle structure. Acetylcholine molecules, on leaving the nerve terminal, cross the primary synaptic cleft (which is continuous with the basal lamina) to activate acetylcholine receptors in the secondary synaptic cleft, consisting of invaginated channels of the nerve fiber's sarcolemma.

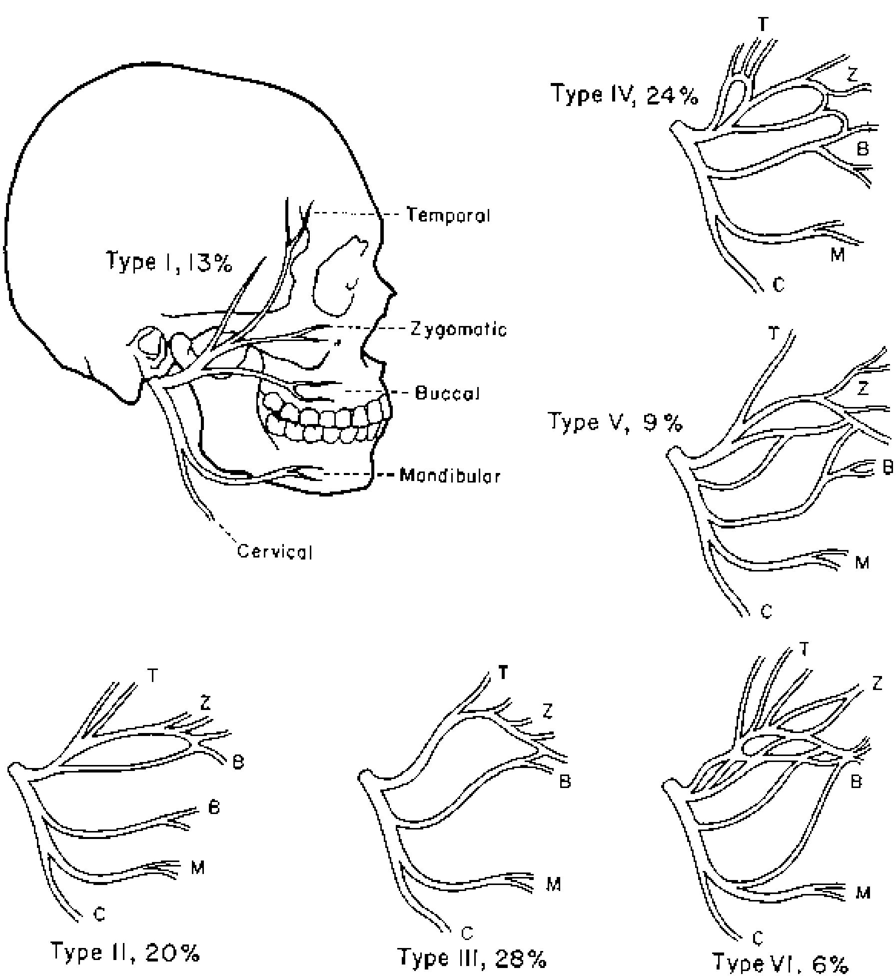


Figure 35–4 *The frequency of natural variations in the extracranial course of the facial nerve. Facial nerve branches: temporal (T), zygomatic (Z), buccal (B), mandibular (M), cervical (C). (From Davis RA, Anson BJ, Puddinger JM, Kurth RE. Surgical anatomy of the facial nerve and parotid gland based upon a study of 350 cervical facial halves. Surg Gynecol Obstet 1956;102:385–412. Reprinted with permission.)*

While acetylcholine molecules travel through the primary synaptic cleft, they are subject to metabolism by acetylcholinesterase. The cholinergic nature of the facial neuromuscular junction forms the basis for botulinum toxin (Botox) treatment of facial muscle hyperfunction (for additional information and applications of Botox, see Chapter 57). Botulinum toxin (trade names Oculinum and Botox) is a neurotoxin produced by *Clostridium botulinum*, a soil anaerobe. This neurotoxin acts by blocking the release of acetylcholine at the nerve terminal.

PATHOPHYSIOLOGY OF THE FACIAL NERVE

Response of the Neuromuscular Unit to Injury

The axon's metabolic demands make it reliant on continuity with the neuron's cell body, its source of nutrients. Likewise, normal structure and function of muscle fibers rely on intact innervation. Electrodiagnositic testing of the facial nerve is based on knowledge of the immediate and delayed events of neuromuscular injury; treatment and prognosis of facial nerve pathology are informed by the results of electrodiagnostic testing.

A stereotypical sequence of events follows interruption of axonal continuity, whether the interruption occurs from compression or transection. Chromatolysis, or dispersal of Nissl substance within the neuron's cell body, is a classic finding first described in the facial nerve nucleus. A general increase in cellular metabolism is reflected in increased numbers of organelles and unregulated oxidative metabolism. Damming of anterograde axoplasmic transport causes the proximal axon stump to swell. Growth cones form within 3 days and extend axon sprouts into the periphery, growing at a rate of ~1 mm per day.

The distal axon stump swells transiently as a result of retrograde axoplasmic damming; within the initial 24 to 48 hours, physiological processes such as axoplasmic transport and nerve conductivity remain largely intact. Thereafter, the effects of wallerian degeneration act to compromise first axoplasmic flow and then myelinization of the affected nerve. Schwann's cells proliferate, some of which perform a phagocytic function and some of which enter endoneural tubes to form Büngner's bands.

Denervation of muscle fibers is manifested as loss of muscle mass and voluntary control. In the initial weeks following such injury, the rate of atrophy is steep but then stabilizes. Fibrillation of the nerve begins within 2 to 3 weeks. With chronic denervation, muscle tissue is eventually replaced by fibrous tissue that contributes to sequelae such as contractures. New motor end plates that are formed after an injury are more numerous than those present before the injury.

Sunderland Classification of Neural Injury

Injury to the facial nerve, commonly caused by inflammatory, neoplastic, or iatrogenic processes, can occur at any point in its course. Sunderland (1978) described five degrees of injury in peripheral nerves (**Fig. 35–5**) The first three degrees of injury can be produced by inflammatory entities such as Bell's palsy and herpes zoster oticus, and the fourth and fifth degrees of injury involve disruption of the nerve as produced by traumatic, neoplastic, or iatrogenic mechanisms. It should be remembered that a single given nerve can exhibit mixed degrees of injury.

First-degree injury, also called neuropraxia, is the loss of conduction across a point of compression or increased intraneural pressure. The nerve is otherwise intact, with no disruption of endoneurium, axons, or axoplasmic flow. With removal of the source of compression, a full return of function occurs.

Second-degree injury, also called axonotmesis, occurs when the endoneurium is intact, but there is axoplasmic damming of nutrients, which results in axonal disruption. This type of injury can occur when there is a more chronic compressive insult to the nerve, and, although the nerve can recover when this compression is relieved, recovery generally takes longer than with neuropraxia. Wallerian degeneration of the distal axon segment occurs.

Third-degree injury, or neurotmesis, results from interruption of endoneurial tubules. When the endoneurium is disrupted, regenerating nerve fibers lose the ability to rejoin corresponding distal segments. During recovery, this cross-wiring of neural projections leads to aberrant regeneration and can cause phenomena such as synkinesis, facial spasm, and gustatory lacrimation. Recovery of full muscle strength is incomplete.

Fourth-degree injury, or partial transection, disrupts the perineurium as well as the endoneurium. The opportunity for regenerating axons to enter incorrect fascicles of a degenerated nerve is therefore increased, and sequelae such as synkinesis are more severe and more frequent. Recovery of full muscle strength is seldom seen.

Fifth-degree injury is complete transection with a disruption of both the epineurium and perineurium. No spontaneous recovery can be expected; therefore, surgical intervention is required. Unrepaired injuries

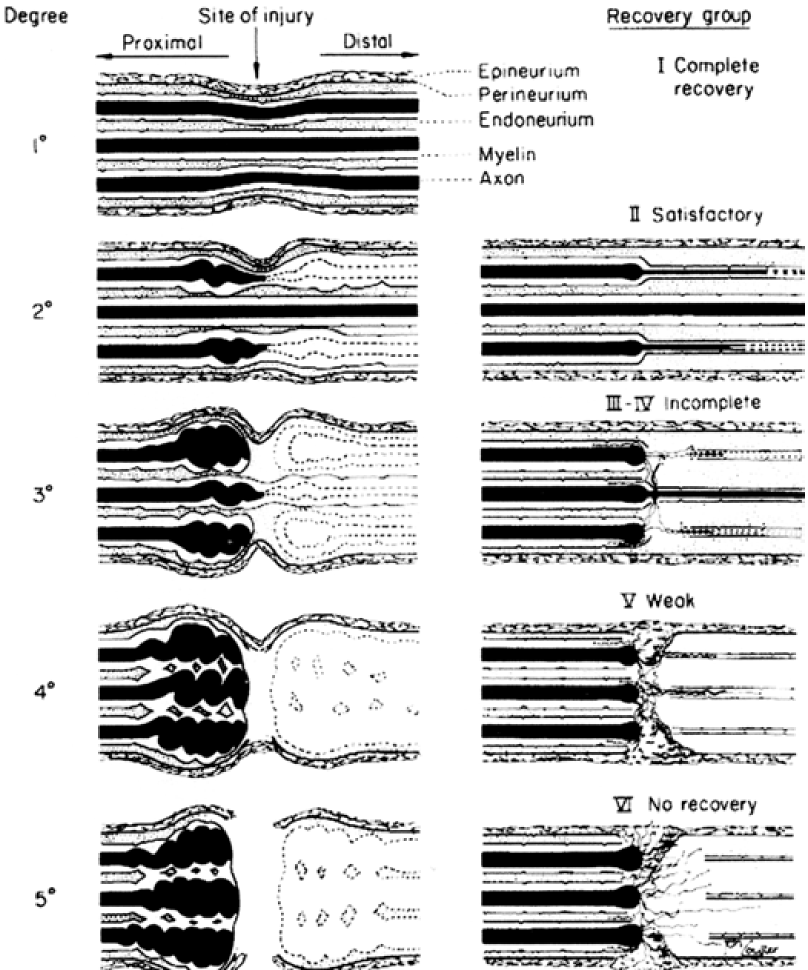


Figure 35–5 *Sunderland classification of the degree of nerve injury and the estimated potential for the recovery of neural function. (From Sunderland, S: Nerve and Nerve Injuries. 2nd ed. London: Churchill Livingstone; 1978:88, 89, 96, 97, 133. Reprinted with permission.)*

exhibit total paralysis of deinnervated areas, and synkinesis is not found.

Aberrant Regeneration

An intact endoneural tube allows a regenerating axon to reestablish communication with the appropriate distal stump. With discontinuity of the endoneural tube, Schwann's cells of Büngner's bands act as a scaffolding for axonal regeneration, but often the proximal axons can reconnect with the wrong distal stump. This "cross-wiring" leads to the phenomena of aberrant regeneration, which include but are not limited to muscle spasticity, synkinesis, gustatory lacrimation, and Frey's syndrome.

ELECTROPHYSIOLOGY

Although the single most important prognostic indicator for the return of facial function after injury is the presence of voluntary movement, numerous electrodiagnostic tests have been developed for the assessment of facial nerve motor function. These tests, especially electroneuronography (ENOG), have largely replaced topognostic testing, which used measures of lacrimation (Schirmer's test) and stapedial reflex to establish the site of lesion. The majority of the electrophysiological assays rely on comparisons between the paralyzed and normal sides of the face and therefore can be inaccurate in the presence of preexisting or congenital facial nerve disorders on the "normal" side.

The nerve excitability test, maximal stimulation test, and ENOG are all evoked response tests, in that they apply a stimulus in the form of a square-wave electrical pulse. Electromyography (EMG), in the setting of facial nerve testing, is a measure of volitional muscle action: the patient is asked to move a muscle group, and resulting motor potentials are recorded. The four tests described above are the most frequently relevant to facial nerve disorders encountered by the otolaryngologist. Although the field of neuromuscular electrophysiology offers other measures of motor nerve function, such as nerve conduction velocity, the majority of facial nerve disorders seen by otolaryngologists are evaluated by the four aforementioned tests.

Nerve Excitability Test

The nerve excitability test (NET) measures the minimal current required to cause a detectable movement in the face. A reference electrode from a stimulator such as the Hilger nerve stimulator is placed between the mastoid tip and the mandible. The stimulating electrode is placed over the approximate location of the pes anserinus. Starting at 0 mA, the threshold current required (normal range 1–4 mA) is found first on the normal side; these thresholds are then compared with the thresholds obtained from the paralyzed side of the face. A difference of 2.0 mA or greater between the normal and paralyzed sides of the face is considered significant.

Maximal Stimulation Test

The maximal stimulation test uses an approach similar to the NET, but grades facial function based on a scale that an observer applies to movements elicited by the highest stimulus intensity the patient can comfortably tolerate (usually ~8–10 mA). Movement of the paretic side of the face is described as normal, slightly decreased, greatly decreased, or absent in comparison to the unaffected (normal) side of the face.

Electroneuronography

ENOG is a measure of a muscle compound action potential (CAP) that uses a paradigm similar to the maximal stimulation test, but that quantifies the response to stimulation. The ENOG employs a stimulating electrode placed between the mastoid tip and the mandible. Maximal stimulus intensity is applied, but it is limited by the patient's pain threshold. This generates a biphasic muscle potential, the sum of many individual motor unit potentials, which is detected by skin surface electrodes. A peak-to-peak amplitude measure of this biphasic CAP is calculated and compared with responses obtained from the intact side (**Fig. 35–6**). This measure, expressed as percent excitability, correlates to the degree of muscle degeneration on the deinnervated side.

Because nerve conduction can remain intact until a sufficient degree of wallerian degeneration has occurred, the ENOG can be unreliable in the initial 48 to 72 hours after a nerve injury. The prognostic utility of ENOG is greatest within 3 weeks of injury; patients manifesting less than 10% excitability within 2 weeks of paralysis onset are classified as having sustained a severe injury. In the face of this degree of injury, the chances of complete recovery are less than 50%, and surgical decompression must be contemplated.

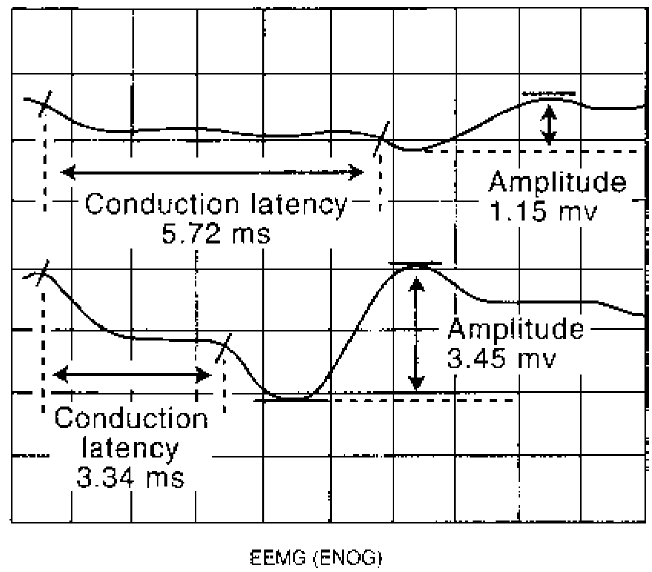


Figure 35–6 *Electroneuronography: muscle compound action potentials from monitoring of patients with a normal facial nerve (bottom trace) and an injured facial nerve (top trace).*

Electromyography

EMG frequently employs needle electrodes to detect striated muscle activity in response to volitional movement. Muscle responses are graded from 1+ to 4+, with 4+ being a normal response and 1+ an absent response. Normal muscle will discharge a short burst of activity with needle insertion, followed by relative electrical silence when the muscle is at rest. Deinnervated muscle exhibits spontaneous fibrillation potentials, a characteristic random firing pattern, when at rest. It will not produce a voluntary motor unit action potential. Fasciculation, a spontaneous firing of an entire motor unit, also can be recorded from deinnervated muscle. Again, the accuracy of EMG is dependent on the time course of wallerian degeneration, and a deinnervated muscle may not exhibit fibrillation potentials until a time period of 48 to 72 hours has elapsed following the initial injury. Reinnervation is heralded by the return of voluntary polyphasic motor unit action potentials.

FACIAL NERVE MONITORING DURING SURGERY

Facial nerve monitoring (FNM) is beneficial during procedures that carry a high risk for facial nerve injury. (e.g., cochlear implantation, revision tympanomastoidectomy, surgery to remove an invasive cholesteatoma, and repair of external auditory canal bony stenosis). FNM is used regularly in otolaryngology, which makes it a state-of-the-art adjunct for advanced ear surgery and lateral skull base surgery. It is also very useful in training centers where residents perform some portions of otologic or parotid operations. However, FNM should never be used as a replacement for anatomical knowledge, technical skill, and clinical judgment of the surgeon.

Physiology

Electrical, mechanical, or thermal stimulation results in depolarization of the facial nerve and a compound muscle action potential (CMAP). This CMAP is the composite electrical activity within the target muscle resulting from synchronous activation of a group of motor neurons. FNM using EMG monitoring essentially measures this electrical activity in the portion of the muscle nearest the recording electrodes and converts the electrical activity to sound via a loudspeaker with or without a visual oscilloscope display of neural activity.

Synchronous activity initiated by electrical stimulation produces a biphasic and well-defined waveform, whereas asynchronous activity related to mechanical stimulation produces a polyphasic pattern. When these FNM response patterns are converted to sound, electrical stimulation results in a well-defined train of pulsed sounds, whereas mechanical stimulation yields a rough burst of acoustic energy.

Electromyography Monitoring

Monitoring of more than one muscle provides additional sensitivity. A typical montage of two-channel bipolar recording includes a pair of needle electrodes in the orbicularis oculi muscle ~3 mm apart and another pair of electrodes in the orbicularis oris muscle. The ground electrode is placed in the musculature of the forehead, and the anode for the monopolar nerve stimulator is inserted at a site in the ipsilateral shoulder.

Neuromuscular blocking agents should not be employed following induction of anesthesia if the facial nerve is to be monitored. It is also crucial that patients be adequately grounded to the monopolar electrocautery unit to permit a safe path for current return. If proper grounding is not achieved, the electrocautery current could potentially find a route through the nerve-monitoring electrodes and result in severe burns to the patient (**Fig. 35–7**).

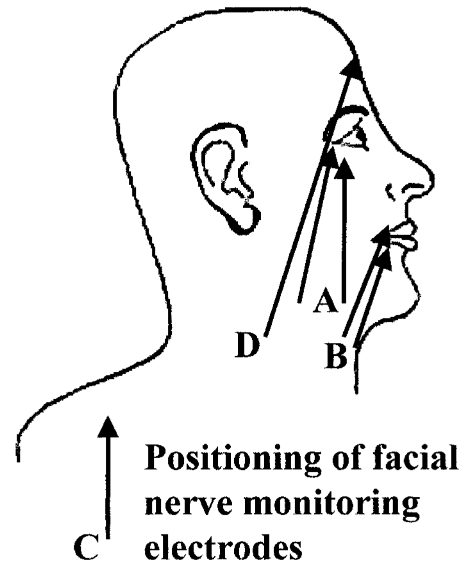


Figure 35–7 *A typical configuration of electrode placement for the two-channel bipolar facial nerve monitoring electrodes. Recording electrodes are placed in two distinct facial muscle groups, indicated by the double arrows in recording sites A and B. The arrows at C and D represent the areas where reference and ground electrodes are placed, respectively.*

Stimulation

The parameters for safe nerve stimulation are 100 to 250 μ sec pulses with a current range of 0.05 to 0.5 mA. Most normal facial nerves should be stimulated by direct contact of the probe using a 1100 μ sec pulse of 0.05 mA. Settings of 0.05 to 0.1 mA are recommended when working close to the nerve. A higher level of stimulation may be required when bone, connective tissue, or granulation tissue covers the nerve.

Burst Activity

Burst activity occurring with gentle manipulation of the nerve is indicative of a healthy facial nerve. During the course of surgery, many burst potentials may be observed, and they are usually not associated with significant trauma to the nerve. Lack of burst activity during dissection may indicate minor manipulation of a healthy nerve, significant manipulation of an already injured nerve, or a problem with the monitoring connections and instrumentation. Electrical stimulation of the nerve at this point can be used to verify the integrity of the nerve or provide an indication of possible damage to this nerve.

Trains of Nerve Responses

Trains of neural discharges are caused by prolonged depolarization of the nerve beyond its threshold for developing an action potential. Subsequent repetitive firing continues until the nerve repolarizes or can no longer sustain repetitive activation. The most common initiating factor for trains of nerve firing is traction on the facial nerve. The occurrence of a train of nerve responses may indicate that significant trauma has occurred to the facial nerve. Changes in temperature around this nerve may precipitate a train of neural activity. When a train pattern of nerve discharge develops after laser application or use of the cautery, thermal damage to the facial nerve should be suspected.

Artifacts

Artifacts are monitoring activities that are not due to facial nerve stimulation. They are a relatively common occurrence during facial nerve monitoring and can cause confusion. Operation of the electrocautery can cause artifacts and obliterate the facial nerve response by saturating the monitor with electrical noise. The operation of an ultrasonic aspirator can cause a large electrical artifact and may trigger a muting circuit. When accurate monitoring cannot be performed, many surgeons engage a member of their operative team to observe visually and/or palpate the patient's face.

Practical Application for Surgery

- Locating the facial nerve and its branches: Monopolar stimulation is especially useful for mapping the facial nerve throughout its course. This is particularly useful for parotid or acoustic neuroma surgeries.
- Avoiding trauma: Continuous EMG monitoring of the facial nerve and its branches provides real-time feedback, enabling the surgeon to alter surgical technique immediately when a burst response occurs.
- Postoperative prognosis: The threshold for stimulation has been shown to correlate well with postoperative facial nerve function.

Suggested Readings

Eshraghi AA, Buchman C, Telischi FF. Sensory auricular branch of the facial nerve. Otol Neurotol 2002;23:393–396

Fisch U, Mattox DE, Valavanis A, Aeppli U. Microsurgery of the Skull Base. New York: Thieme Medical Publishers; 1988

May M, ed. The Facial Nerve. May's 2nd ed. New York: Thieme Medical Publishers; 1986

Schuknecht HF, Gulya AJ. Anatomy of the Temporal Bone with Surgical Implications. Philadelphia: Lea & Febiger; 1986

Sunderland S. Nerve and Nerve Injuries. 2nd ed. London: Churchhill Livingstone; 1978

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. An early indicator of the return of facial nerve function is the
   A. Summating potential
   B. Polyphasic action potential
   C. Relative electrical silence
   D. Return of evoked motor potentials
2. The nervus intermedius (of Wrisberg) carries
   A. Special visceral efferent fibers, whose cell bodies are located in the geniculate ganglion, and which project to the nucleus tractus solitarius
   B. General sensory afferents terminating in the spinal trigeminal tract
   C. Parasympathetic general visceral efferents to the salivary and lacrimal glands, as well as special visceral afferent fibers from the chorda and general sensory efferents from the external auditory canal and postauricular area
   D. Corticobulbar fibers originating in the somatomotor cortex
3. Neurotmesis
   A. Disrupts perineurium as well as endoneurium
   B. Does not disrupt endoneurium, axons, or axoplasmic flow; the nerve is intact
   C. Would not be expected to exhibit any signs of recovery
   D. Results from interruption of endoneurial tubules
4. In the internal auditory canal, the vertical crest (Bill's bar) marks a plane between
   A. Superior and inferior vestibular nerves
   B. Facial and cochlear nerves

   C. Inferior vestibular and singular nerves
   D. Superior vestibular and facial nerves

5. The nerve excitability test
   A. Is an assessment of the minimal current required to evoke a facial movement
   B. Is a subjective comparison of the degree of facial movement in the paralyzed side versus the intact side
   C. Grades voluntary potentials from 1+ to 4+, with 4+ being a normal response and 1+ an absent response
   D. Is most accurate within 48 hours of an injury

6. Aberrant regeneration
   A. Can lead to gustatory lacrimation
   B. Has some symptoms that can be treated with injections of botulinum toxin (Botox)
   C. Is a signficant cause of postparalysis morbidity
   D. All of the above

7. The tympanic segment of the facial nerve
   A. Originates immediately distal to the meatal foramen and is the shortest segment, 3 to 4 mm in length
   B. Includes a portion that, at 0.68 mm average diameter, is the narrowest part of the facial nerve
   C. Maintains a superior relationship to the canal of the tensor tympani muscle and cochleariform process
   D. Is separated from the superior vestibular nerve by Bill's bar

8. The single most important prognostic factor for eventual recovery of facial nerve function is
   A. Presence of voluntary movement
   B. Less than a 2.0 mA difference on nerve excitability testing, relative to the unparalyzed side
   C. Greater than a 2.0 mA difference on nerve excitability testing, relative to the unparalyzed side
   D. An intact stapedius reflex on topognostic testing

9. Facial nerve monitoring may be beneficial during the following procedure(s):
   A. Revision tympanomastoidectomy
   B. Cochlear implantation
   C. Acoustic neuroma surgery
   D. All of the above

Chapter 36

RADIOLOGY OF THE TEMPORAL BONE

Barbara Zeifer

Temporal bone imaging is a daunting topic for both radiology and otolaryngology residents. For the radiologist, the ear is a difficult region because of the complex and unfamiliar anatomy. For the otolaryngologist, the difficulty lies in the unfamiliar imaging planes: otoscopy and surgery are approached from a lateral view, whereas the images provided are generally those in the axial and coronal projection. The otolaryngologist must therefore "translate" the familiar anatomy into a less familiar view.

This chapter eases the transition from direct visualization to cross-sectional imaging, beginning with a discussion of normal anatomy. This is followed by a review of inflammatory disease of the mastoid and middle ear, trauma, vascular lesions of the middle ear, lesions of the petrous apex, and lesions of the cerebellopontine angle cistern, and concludes with a brief discussion of congenital malformations of the inner and external ear.

IMAGING TECHNIQUES

Plain radiography was originally felt to provide a "screening" examination of the entire temporal bone. Seven conventional projections were available: Law, Schüller, Owen, Chausse III, Towne, Stenvers, and transorbital views were used in various combinations to demonstrate the different areas of the temporal bone. As with all plain imaging studies, only gross bone detail could be seen. Mastoid pneumatization could be assessed, pronounced anomalies could potentially be detected, but soft tissue pathology would be apparent only if osseous structures were eroded severely enough to be detected on the x-ray examination. Most radiology departments no longer maintain the specialized head units necessary to produce high-quality plain films of the temporal bone, and it is a rare technologist who remembers how to perform the proper views. Plain film imaging of the temporal bone has now been completely replaced by computed tomography (CT) and magnetic resonance imaging (MRI).

Computed tomography offers magnificent detail of bony structures. It is the imaging modality of choice in the evaluation of the osseous external canal, mastoid air cells, middle ear and ossicles, course and caliber of the facial nerve, and otic capsule. Images are generated by detection of a transmitted x-ray beam, and tissue contrast results from differences in electron density. We refer to the density of various tissues on a CT scan: higher density tissues are whiter, and lower density tissues are darker. Bone and calcifications are white, air is black, and for temporal bone imaging, everything else (i.e., fluid, blood, debris, and cellular tissue) is gray. The narrow slice thickness (1.0–1.5 mm) and the algorithms used to produce the image ("bone, "sharp," or "edge") make it impossible to further discriminate between these substances.

Intravenous contrast in temporal bone CT is indicated in three situations. The first is to evaluate a patient for asymmetrical sensorineural hearing loss and acoustic neuroma who cannot have an MRI, the second is to evaluate the vascular nature of a middle ear mass, and the third is to evaluate the jugular bulb for tumor invasion.

As does CT, MRI provides cross-sectional images of the body. Simplistically, the MR image is produced by detection and measurement of emitted energy following the rapid pulsation of a radiofrequency wave into a high-strength magnetic field (up to 1.5 tesla). Therefore, we refer to the signal intensity of various tissues on the MRI scan: hyperintense tissues are brighter, and hypointense tissues are darker. Four physical properties of matter are responsible for production of signal intensity in MRI: the T1 and T2 relaxation times, proton density, and flow. Data can be generated in different ways by varying the rapidity of the radiofrequency excitations, the direction of magnetic gradients, and the timing of data measurement to produce different sequences. Standard sequences include T1-weighted (T1W), proton density or balanced, and T2-weighted (T2W) sequences. It is helpful to remember that water (e.g., cerebrospinal fluid [CSF] and cysts) is hypointense on T1W and hyperintense on T2W images, whereas fat is hyperintense on T1W and fades on T2W. Both cortical bone and air generate no signal and are completely black, or signal void; medullary bone, however, contains fatty marrow and will be bright on T1W sequences.

The weakness of MRI in temporal bone imaging is the lower spatial resolution compared with CT. The strength of MRI lies in its ability to produce soft tissue detail and tissue characterization. A lesion that is of nonspecific soft tissue density on CT will often emit a pattern of signal intensities on MRI that will enable a specific diagnosis to be made. MRI is the imaging modality of choice in the evaluation of the internal auditory canal (IAC) contents, the cerebellopontine angle (CPA) cistern, brainstem, cerebellar hemispheres, and membranous labyrinth. It complements CT in the evaluation of skull base pathology.

Intravenous contrast is extremely useful in MRI. In T1W imaging there is less natural contrast between the otic capsule, membranous labyrinth, nerves, and CSF. Following contrast injection, however, abnormal enhancement of the membranous labyrinth and neural elements of the IAC will be clearly visualized. Contrast is also useful in detecting intracanalicular acoustic neuromas that are 3 mm or less in diameter; larger tumors can be identified without contrast administration but are more conspicuous following the injection. The presence or lack of contrast enhancement will aid in refining the differential diagnosis of mass lesions.

Magnetic resonance angiography (MRA) and magnetic resonance venography (MRV) are additional tools available for use in the evaluation of the posterior fossa, but they can occasionally be misleading. Time-of-flight angiography is performed routinely and demonstrates high-flow vascular structures with excellent spatial resolution. Phase-contrast angiography images flow at a specific velocity selected by the operator during setup; for example, selecting 80 cm per sec demonstrates vessels with high flow, while selecting 5 cm per sec demonstrates structures with slow-flow (i.e., veins). Phase-contrast angiography is therefore extremely useful for venography. Routine MRA images of blood flow require laminar flow to generate a signal that can be imaged. This type of examination is subject to a wide range of artifacts generated by various flow phenomena. Imaging flow is distinctly different from imaging the structure of a vessel lumen, as in catheter angiography, and cannot be interpreted in the same way. It is possible to have a "normal" MRA/MRV with a glomus tympanicum that has begun to invade the jugular bulb, as well as with an aneurysm that contains hyperacute thrombus. Subacute hemorrhage or thrombosis is high in signal on time-of-flight angiography, so that a thrombosed vessel could appear to have normal flow. Clearly, then, MRA/MRV is not a gold standard for vascular imaging and must be interpreted with a good-quality MRI as well as clinical input.

NORMAL ANATOMY

Thin-section, high-resolution images are generated in a routine temporal bone CT study. Both axial and direct coronal scans are performed and provide complementary information. Various structures are visualized depending on their shape and orientation in relation to the slice plane.

Evaluation of the axial CT (**Fig. 36–1**) begins inferiorly, with identification of the mastoid air cell system,

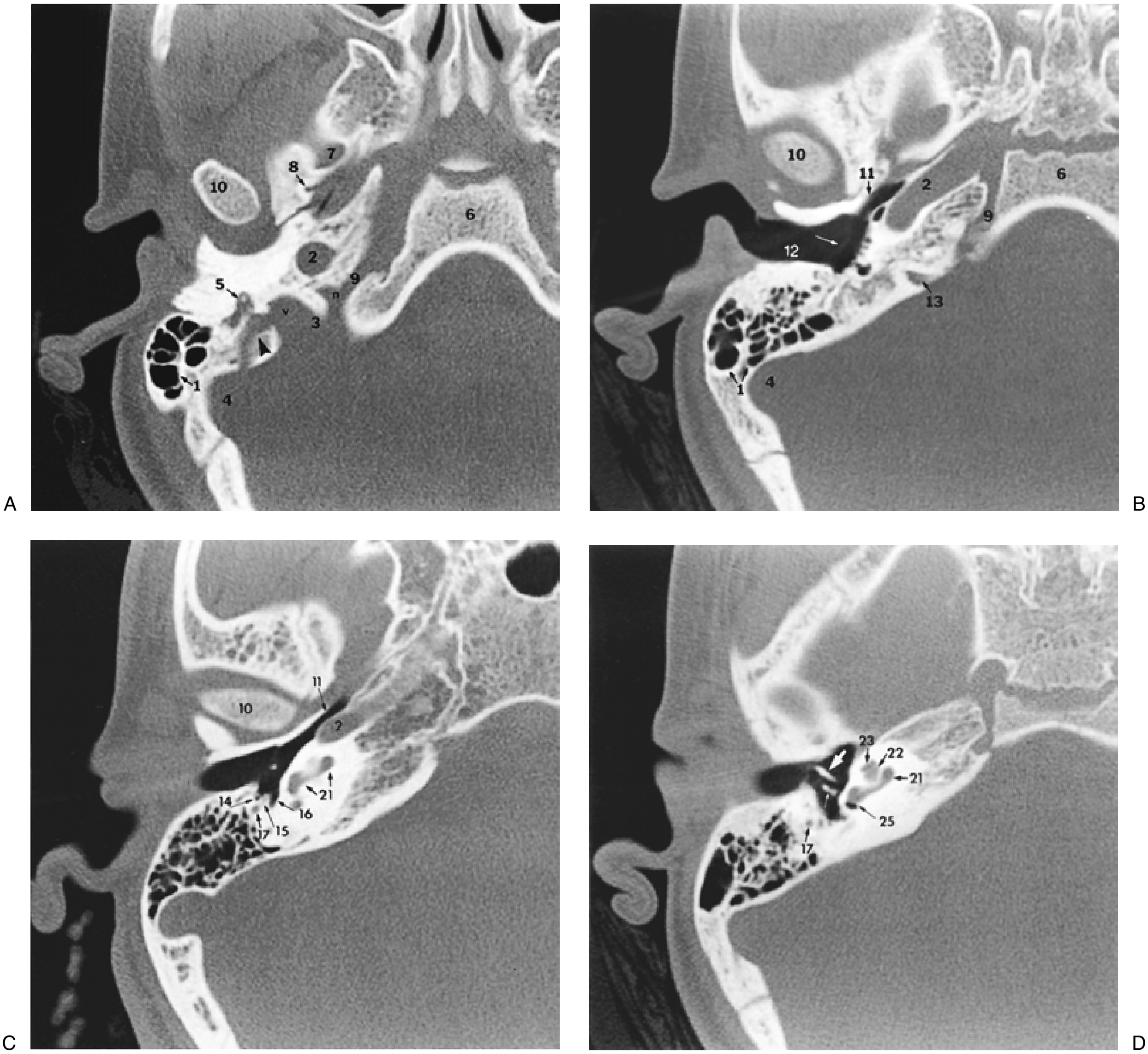
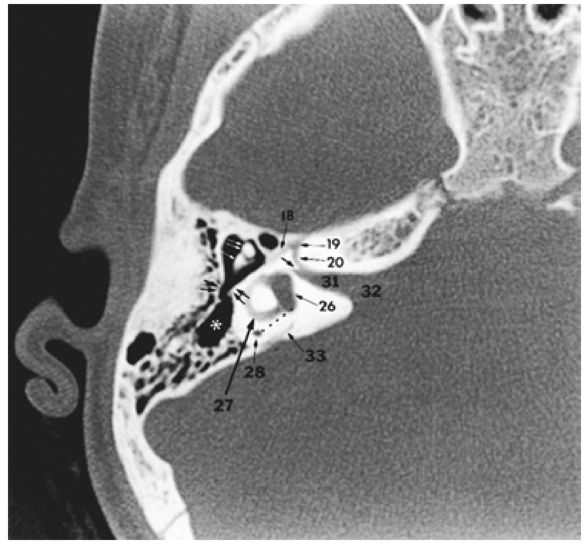


Figure 36–1 *Normal anatomy, axial computed tomography. Images progress from inferior* ***(A)*** *to superior* ***(G)****. Images are from different patients.* ***(A)*** *Skull base. Skull base foramina are well delineated. An unusual emissary vein drains into the jugular bulb (arrowhead).* ***(B)*** *Eustachian tube. The bony eustachian tube lies lateral and anterior to the horizontal segment of the petrous carotid canal (white arrow, tympanic membrane).* ***(C)*** *Posterior tympanic recesses.* ***(D)*** *Ossicles and inner ear: mesotympanum (thick white arrow, manubrium of malleus; thin white arrow, lucent space of incudostapedial joint). Medial to the joint space (toward the cochlea) is the head of the stapes; lateral to the joint is the lenticular process of incus.* ***(E)*** *Ossicles and inner ear: mesotympanum (double white arrow, neck of malleus; single white arrow, long process of incus; black arrows point to margins of the oval window). The stapes crura are seen as thin white linear structures projecting into the oval window; the arrow pointing to the modiolus lies in the fundus of the internal auditory canal.* ***(F)*** *Ossicles and inner ear: epitympanum. Malleus and incus have an ice cream cone appearance (double white arrow, malleus head; single white arrow, body of incus; the curved black line between the two is the malleoincudal articulation; double black arrows, aditus ad antrum; asterisk, mastoid antrum; single black arrow, Bill's bar). Note the wispy configuration of the vestibular aqueduct, running roughly parallel to the axis of the posterior semicircular canal (dotted line).* ***(G)*** *Körner's septum (asterisk, mastoid antrum; double arrow, Körner's septum). 1, mastoid air cells; 2, carotid canal; 3, jugular foramen; 4, sigmoid sinus; 5, stylomastoid foramen; 6, clivus; 7, foramen ovale; 8, foramen spinosum; 9, petro-occipital fissure; 10, mandibular condyle; 11, bony eustachian tube; 12, external auditory canal; 13, cochlear aqueduct; 14, facial recess; 15, pyramidal eminence with stapedius muscle; 16, sinus tympani; 17, descending facial nerve canal; 18, horizontal facial nerve canal; 19, geniculate ganglion; 20, labyrinthine facial nerve canal; 21, basal turn of cochlea; 22, middle turn of cochlea; 23, apical turn of cochlea; 24, modiolus; 25, round window; 26, vestibule; 27, lateral semicircular canal; 28, posterior semicircular canal; 29, superior semicircular canal; 30, crus commun; 31, internal auditory canal; 32, porus acusticus; 33, vestibular aqueduct.*

E



F

G

Figure 36–1 *(Continued)*

the bony external auditory canal (EAC), and the skull base foramen. Anteriorly, the hypotympanum narrows as it becomes confluent with the eustachian tube. The tympanic membrane (TM) may be seen as a wisp of tissue traversing the tympanic ring; medial to the TM lies the middle ear cavity.

To evaluate the ossicular chain, begin superiorly within the epitympanum, above the tympanic ring. At this level, the malleus and incus will have the appearance of an ice cream cone: the head of the malleus is the ice cream ball, articulating with the body of the incus as the ice cream cone. The malleoincudal articulation is seen in the axial view. The tip of the ice cream cone is the short incudal process that projects into the fossa incudis. As one proceeds inferiorly into the mesotympanum, the head of the malleus tapers to form the neck and, more inferiorly, the manubrium, which is embedded in the TM; the body of the incus tapers to the long process. The stapes lies in the horizontal plane and therefore will be most clearly imaged on the axial slices, with the crura extending medially into the oval window. A thin white line traverses the oval window, representing the stapes footplate.

The posterior tympanic recesses are also well seen in the axial plane. Medially, the sinus tympani is of variable depth. Lateral to the sinus tympani is the pyramidal eminence, out of which the stapedius tendon extends. The muscle is often seen within the bone at the center of the eminence, but a normal stapedius tendon is not visible on CT. Lateral to the pyramidal eminence is the shallow facial recess.

The epitympanum opens posteriorly into the mastoid antrum via the hourglass-shaped aditus ad antrum. The antrum is variable in size. Körner's septum, the inferior extension of the petrosquamosal suture, can be found laterally within the antrum as an osseous bar running parallel to the long axis of the petrous bone.

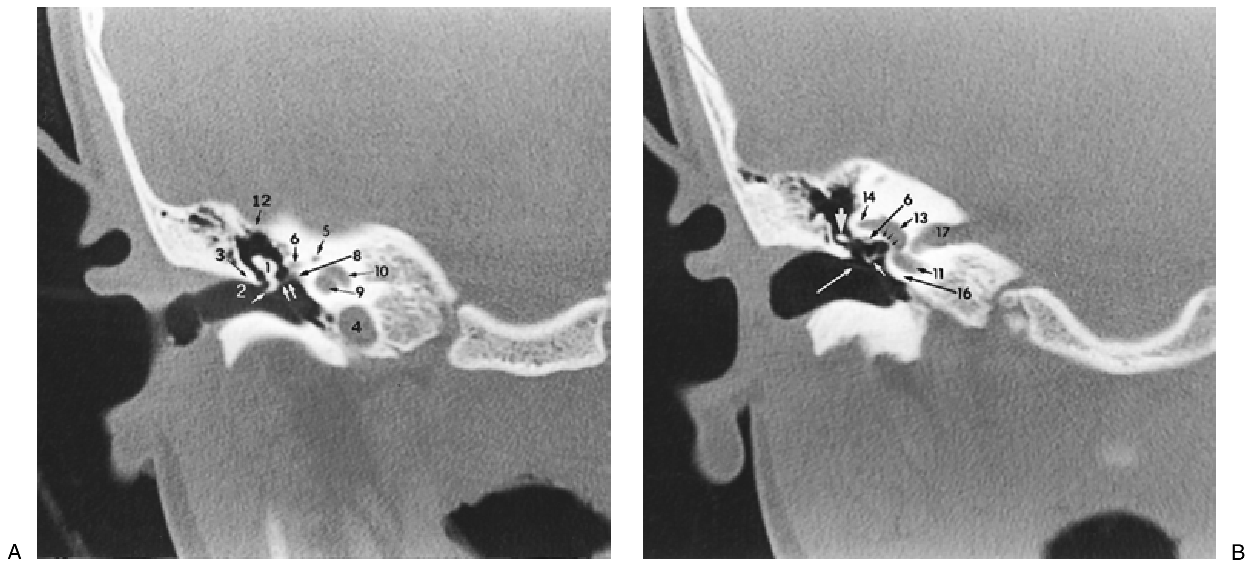
Evaluation of the inner ear structures begins by finding the internal auditory canal as a prominent lateral extension of the posterior fossa into the temporal bone, ~4 to 5 mm in diameter. The lateral wall of the IAC is the lamina cribrosa, and the superior fundus is divided vertically by Bill's bar, a triangular strut of bone seen only on axial images. The medial opening of the IAC to the posterior fossa is the porus acusticus. Anterior to the fundus of the IAC lies the cochlea, with its basal, middle, and apical turns. The round window is found along the basal turn, posteriorly, behind a bony prominence, or hook. The central axis of the cochlea, the modiolus, is contiguous with the IAC and anchors the osseous spiral lamina. These septations produce a faint *Y* or *V* bony density on axial CT; the lateral septation is often faint. Posterior and lateral to the IAC and posterior to the cochlea lies the vestibule. The lateral semicircular canal (LSCC) projects laterally from the vestibule and will often be seen in its entirety on the axial scan. The superior semicircular canal (SSCC) projects superiorly; its anterior and posterior rims will be cut in cross section until the apex, where they become confluent. The posterior semicircular canal (PSCC) projects posteriorly from the vestibule. The posterior rim of this canal will be cut in cross section centrally; inferiorly it becomes confluent with its anterior rim, and superiorly, joins the posterior portion of the SSCC to form the thick common crus.

Both of the aqueducts are found on the axial CT scan; each acts as a conduit between the membranous labyrinth and the subarachnoid space of the posterior fossa. The landmark for the vestibular aqueduct (VA) is the axis of the PSCC; the VA runs nearly parallel and just posterior to the PSCC, sweeping from the endolymphatic sac (not seen on any imaging study) to the vestibule. The normal VA can be as thick as 2 mm in diameter but may not be identified at all; it can be straight or S-shaped. The landmark for the cochlear aqueduct (CA) is the jugular bulb. Anterior and superior to the jugular canal a triangular extension of posterior fossa CSF will be identified. This portion of the CA may be variable in size and is often asymmetrical. A thin, straight, lucent line extending from the cranial orifice of the CA to the basal turn of the cochlea represents the intraosseous segment of the aqueduct, measuring less than 1 mm in diameter.

Evaluation of the coronal CT (**Fig. 36–2**) begins with identification of the EAC, and the thin TM traversing the tympanic ring; the scutum is a triangular strut of bone that provides the attachment for the superior TM. The malleus is the most anterior ossicle: the head, neck, lateral short process, and manubrium can all be seen. Prussak's space is identified at this anterior level and is bounded laterally by the scutum, medially by the neck of the malleus, inferiorly by the lateral short process of the malleus and the pars flaccida segment of the TM, and superiorly by the lateral suspensory ligament of the malleus, which is not normally visualized on CT.

The tensor tympani muscle originates from the superior margin of the bony eustachian tube, then runs posteriorly along the medial wall of the tympanic cavity immediately below the proximal tympanic segment of the facial nerve canal. It can be identified as a tissue band, inferior to the horizontal facial nerve (HFN). The muscle makes a right-angle turn at the cochleariform process, a lateral projection of the semicanal in which the tensor tympani runs. The tendon then sweeps laterally to insert onto the neck of the malleus and is visualized as a thin tissue band. Immediately behind the malleus lies the incus: the body will be cut cross section and rounded, while the long process extends inferiorly in a smooth, gradual, medial-convex arc. The long process of the incus terminates in the small lenticular process with a 90-degree angulation; the lenticular process projects medially and articulates with the stapes capitulum. The incudostapedial articulation is seen in the coronal view as a faint lucency between the two ossicles. One or both stapedial crus can be seen running medially into the oval window niche, a rectangular "recess" of the middle ear cavity that is best appreciated in the coronal plane. The oval window is traversed by the stapes footplate and forms the junction between the inner and middle ear.

The cochlea is found on the most anterior slices of the coronal CT as a coiled structure with its apex directed anterolaterally. The osseous spiral lamina should be seen to extend into the apex of the cochlea, between the middle and apical turns (see **Fig. 36–3**). The bone over the basal turn bulges laterally into the mesotympanum, can be seen on otoscopy, and represents the cochlear promontory. The most posterior slice through the cochlea demonstrates the basal turn below and becoming confluent with the vestibule. It is important not to confuse this portion of the cochlea with the thinner semicircular canals on the coronal CT. The IAC is seen medially, its fundus divided horizontally by the crista falciformis (see **Fig. 36–4**). Lateral to the lamina cribrosa is the vestibule, and along the lateral wall of the vestibule lies the oval window. The SSCC projects superiorly from the vestibule; the LSCC projects



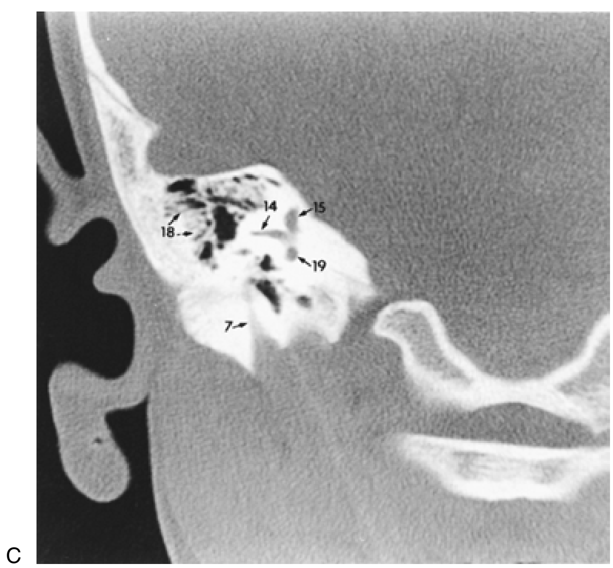


Figure 36–2 *Normal anatomy, coronal computed tomography. Images progress from anterior **(A)** to posterior **(C)**. **(A)** Malleus (double arrows, tensor tympani tendon; single arrow, lateral short process malleus). Prussak's space is well delineated. **(B)** Incus and stapes (large white arrow, body incus; small white arrow, incudostapedial joint: lateral to the joint is the lenticular process of incus; medial to the joint is the stapes capitulum; short black arrows, oval window; the stapes crura are well seen projecting into the oval window; long white arrow, tympanic membrane). The facial nerve canal has completely formed. **(C)** Posterior inner ear. The posterior rim of the lateral and superior semicircular canals are seen with the anterior rim of the posterior canal. 1, head of malleus; 2, external auditory canal; 3, scutum; 4, carotid canal; 5, labyrinthine facial nerve canal; 6, tympanic facial nerve canal; 7, descending facial nerve canal; 8, tensor tympani muscle at cochleariform process; 9, apical turn cochlea; 10, middle turn cochlea; 11, basal turn cochlea; 12, tegmen tympani; 13, vestibule; 14, lateral semicircular canal; 15, superior semicircular canal; 16, cochlear promontory; 17, internal auditory canal; 18, mastoid air cells; 19, posterior semicircular canal.*

laterally; the PSCC projects posteriorly. The bony encasement of the LSCC forms the roof of the oval window niche.

The radiographic anatomy of the facial nerve (FN) is complex because its convolutions run in and out of the scan plane. On axial CT, the labyrinthine segment of the FN canal originates from the anterior superior quadrant of the IAC. The nerve runs anteriorly as a curved lucency within the otic capsule and proceeds to the geniculate ganglion. From there the FN makes a "U-turn" (anterior genu) and runs posteriorly in the medial wall of the middle ear cavity, above and parallel to the tensor tympani muscle. The FN continues posteriorly below the LSCC, and after emerging from the posterior rim, makes a 90-degree turn inferiorly (posterior genu). The posterior genu is located just above the pyramidal eminence, after which the descending or mastoid segment of the nerve runs vertically downward behind the facial recess and pyramidal eminence. It continues inferiorly to exit the stylomastoid foramen.

On the coronal scans, the facial nerve ganglion is a rounded lucency within the bone on the most anterior slices. Behind the ganglion and anterior genu, the nerve "branches" into the labyrinthine segment medially and the tympanic segment laterally, producing two rounded lucent dots lateral and superior to the middle and apical cochlear turns. The labyrinthine nerve runs into the IAC, and the tympanic segment runs posteriorly in the medial wall of the middle ear. The coronal view is optimal for visualizing the horizontal segment of the FN canal because the canal is "cut" in cross

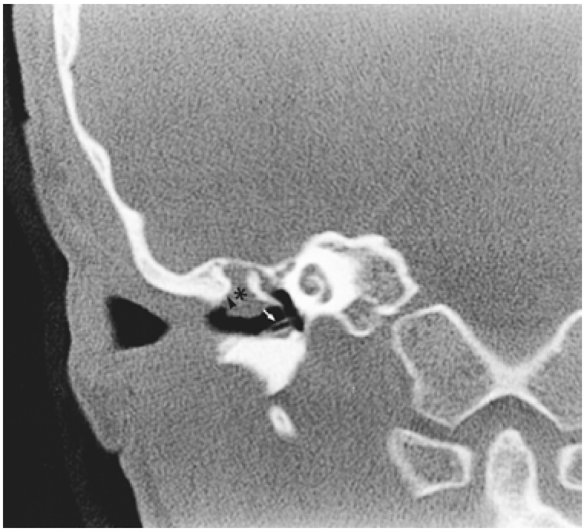


Figure 36–3 *Attic cholesteatoma, coronal computed tomography. There is a bulky soft tissue mass in Prussak's space (asterisk). The scutum (arrowhead) and lateral short process of malleus are eroded. Prussak's space is widened, and the malleus is displaced medially. A tympanostomy tube is in place (white arrow). Note the normal extension of the osseous spiral lamina into the apex of the cochlea.*

section and is clearly identified just below the LSCC, lateral and superior to the oval window. The inferior wall of the horizontal facial nerve canal is often thin and not well visualized. If a notch or divot can be seen in the bone above the nerve, then the canal has completely formed, and congenital dehiscence is unlikely. After emerging from the posterior rim of the LSCC, the nerve makes its posterior genu, then descends as the mastoid segment. This segment may be seen on one to three images, depending on its orientation to the vertical slice plane.

Less anatomical detail is present on MRI of the temporal bone as compared with CT. Wider slices result in fewer images, and the lack of air–bone contrast (both are signal void) limits visualization of the middle ear. On high-resolution T2W sequences through the IAC and CPA, it is possible to delineate the neural structures as distinct from CSF because the fluid is bright and the nerves are low in signal intensity (**Fig. 36–5**). Portions of the membranous cochlea, vestibule, and semicircular canals can be seen due to the high signal intensity of their fluid contents, whereas the surrounding otic capsule is dark and featureless. The carotid canal and jugular foramen can be identified. The CA is a triangular shaped CSF-filled tract just anterior and superior to the jugular bulb. The middle ear contour and contents are not visualized at all. MRI is the imaging modality of choice to evaluate the posterior fossa,

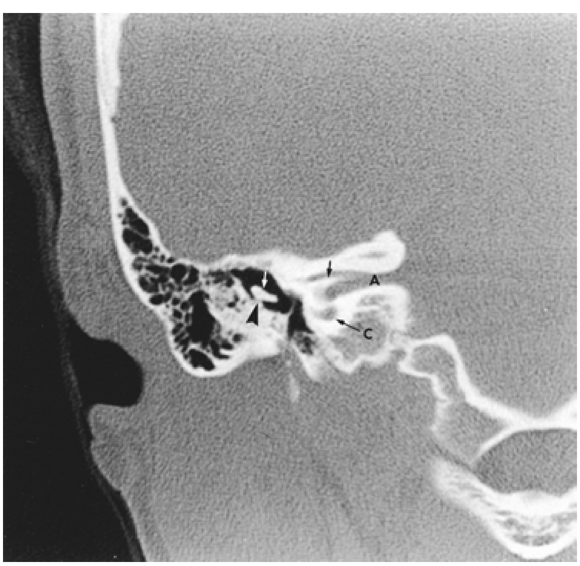


Figure 36–4 *Microtia and aural atresia, coronal computed tomography. The external auditory canal has not formed. In its place are mastoid air cells. The middle ear cavity is moderately narrow. Anteriorly, the malleus and incus are clumped (white arrow) and fused to the adjacent atresia plate (arrowhead). The crista falciformis (black arrow) is prominent. C, cochlea; A, internal auditory canal.*

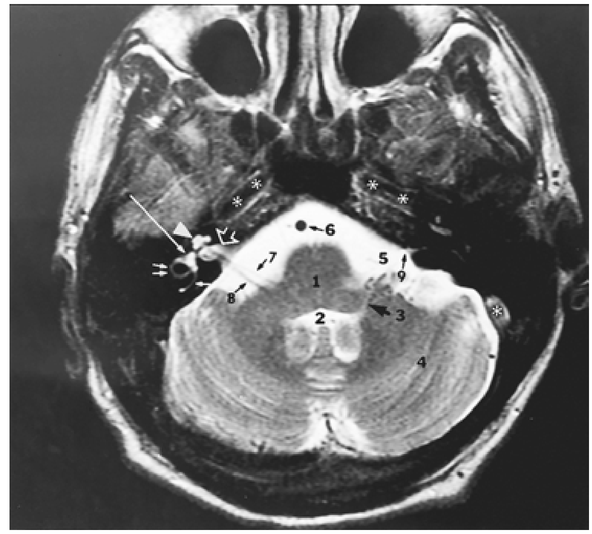


Figure 36–5 *Normal anatomy, axial T2-weighted magnetic resonance image through the inferior portion of the internal auditory canal. 1, brainstem at junction of pons and medulla; 2, fourth ventricle; 3, lateral recess fourth ventricle (foramen of Luschka); 4, cerebellar hemisphere; 5, cerebellopontine angle cistern; 6, basilar artery; 7, facial nerve, cisternal segment; 8, vestibulocochlear nerve, cisternal segment; 9, cochlear aqueduct; single asterisk, sigmoid sinus; double asterisk, carotid canal; open arrow, internal auditory canal with cochlear and inferior vestibular nerves; arrowhead, cochlea (apical and middle turns); long arrow, vestibule; double arrow, lateral semicircular canal; single short arrow, posterior semicircular canal.*

where brainstem and cerebellar tissue is clearly visualized. On T2W sequences, the vertebral-basilar arterial system is seen as dark flow voids surrounded by high-signal CSF.

INFLAMMATORY DISEASE

Without exception, inflammatory disease of the middle ear should be evaluated with CT. Contrast administration offers no additional information and should not be injected. As mentioned earlier, CT offers limited tissue and fluid characterization: we can identify opacification of the middle ear cavity but cannot be more specific. Fluid, blood, pus, thickened mucosa, cholesteatoma, and even tumor have the same density. Therefore, interpretation of the CT scan requires analysis of disease location and the effect this disease produces on middle ear structures. It is this information that will often lead to the diagnosis.

In serous and acute otitis media, the middle ear will be opacified without any abnormality of surrounding bone or ossicles. An air/fluid level that shifts in location from axial to coronal positioning may be present. With chronic otitis media (COM), there is nondependent, usually incomplete opacification of the middle ear cavity. A thickened and/or perforated TM may be seen. In most cases, the ossicular chain will be intact. Occasionally, COM can affect the ossicular chain. Most commonly, demineralization and erosion of the distal incudal body occur and can be extensive. Fixation of the incudostapedial articulation can occur with widening of the joint space due to deposition of fibrous tissue. Erosion of the malleus and body of the incus is uncommon. It should be clear that an opacified middle ear without ossicular involvement could represent serous, acute, or chronic otitis media; the distinction between these entities will be made clinically.

Cholesteatoma has a specific pattern of destruction that is recognizable on the CT scan. A small or early cholesteatoma can of course be missed when imaged prior to the development of these changes. The most common cholesteatoma, the attic cholesteatoma (**Fig. 36–3**), arises from the thin, pars flaccida segment of the TM and grows into Prussak's space. As it enlarges, the cholesteatoma erodes the scutum and then displaces the malleus medially, thus widening Prussak's space. The head, neck, and lateral short process of the malleus and eventually the body of the incus are destroyed. The less common cholesteatoma arises from the pars tensa portion of the TM and has a different sequence of destruction. This lesion grows first into the mesotympanum, particularly the facial recess and sinus tympani, and is therefore called a sinus cholesteatoma. The sinus cholesteatoma tends to erode the long incudal process first, displacing the remainder of the ossicular chain laterally, narrowing Prussak's space. The body of the incus is destroyed next, followed by the malleus. Aggressive cholesteatomas can destroy the otic capsule and erode into the inner ear. The LSCC is the most common site for these labyrinthine fistulas (**Fig. 36–6**).

Figure 36–6 *Sinus cholesteatoma, coronal CT. There is a bulky mass lesion within the mesotympanum and epitympanum. Most of the incus and stapes are eroded; a remnant of incus is displaced laterally (arrow), narrowing Prussak's space. The cholesteatoma has eroded into the lateral semicircular canal to produce a fistula. This was no surprise, however, because the patient was vertiginous on clinical exam when the cholesteatoma was manipulated. Extreme thinning and dehiscence of the tegmen tympani are also seen. The tympanic segment of the facial nerve canal is congenitally dehiscent, and the exact location of the nerve cannot be determined.*

Congenital cholesteatoma accounts for 2% of all cholesteatomas and represents an epidermoid cyst. This lesion typically arises anteriorly and superiorly in the middle ear and results in bony and ossicular erosion that reflects the site of origin. Associated eustachian tube obstruction is common, and the retained secretions surround and obscure the lesion on the CT scan.

Varying degrees of mastoid involvement accompany middle ear disease. In acute or serous otitis, the mastoid air cell system is appropriately pneumatized for the patient's age but is opacified. This opacification simply indicates eustachian tube dysfunction with retained secretions. In the adult, chronic otitis media is associated with thickened osseous septations. Eventually, the mastoid bone becomes completely sclerotic. In children with chronic otitis media, the mastoid never pneumatizes. The long-standing eustachian tube obstruction that leads to the development of cholesteatoma is invariably associated with a sclerotic or nonpneumatized mastoid; it is possible but rare to find a patient with cholesteatoma and a well-pneumatized mastoid bone. True mastoiditis is due to a bacterial infection, and the diagnosis should not be made on CT in the absence of the characteristic clinical features. Initially, mastoiditis is seen as varying degrees of mastoid air cell opacification, often with air/fluid levels, and at this point is radiographically indistinguishable from eustachian tube dysfunction. Mastoiditis can be complicated by loss of septations due to necrosis (osteomyelitis), with coalescence of air cells (coalescent mastoiditis), and eventual cortical destruction.

TEMPORAL BONE TRAUMA

Temporal bone fracture usually occurs in the setting of severe head trauma, and neurological injury must be assessed first, preferably with MRI. Radiographic evaluation of temporal bone fracture and ossicular disruption requires the use of CT. If a fracture is found to extend into the petrous carotid canal, MRA or catheter angiography must be performed to exclude dissection and pseudoaneurysm. If a sensorineural hearing loss is diagnosed and the CT is normal, then an MRI could be performed to detect hemorrhage into the membranous labyrinth, identified as high signal on the noncontrast T1W scan.

Most fractures of the temporal bone (70–90%) are oriented along the long axis of the petrous bone and are referred to as longitudinal fractures (**Fig. 36–7**). Approximately 10 to 20% of longitudinal fractures will involve the facial nerve; many will involve the ossicular chain. Less common is the transverse fracture, oriented perpendicular to the long axis of the petrous bone, half of which will have facial nerve involvement (**Fig. 36–8**). Many temporal bone fractures will be "mixed." Regardless of fracture orientation, the images must be fully assessed for location and extent of injury. True fractures must be distinguished from normal structures that can simulate bone disruption, such as the petromastoid suture, subarcuate canal, cochlear and vestibular aqueducts, posterior ampullary canal, mastoid canaliculus, and inferior tympanic canaliculus.

When conductive hearing loss is diagnosed in the setting of head trauma, disruption of the ossicular chain or oval window has occurred. The least stable ossicle is the incus, which lacks significant attachments. Incudostapedial dislocation is most commonly encountered.

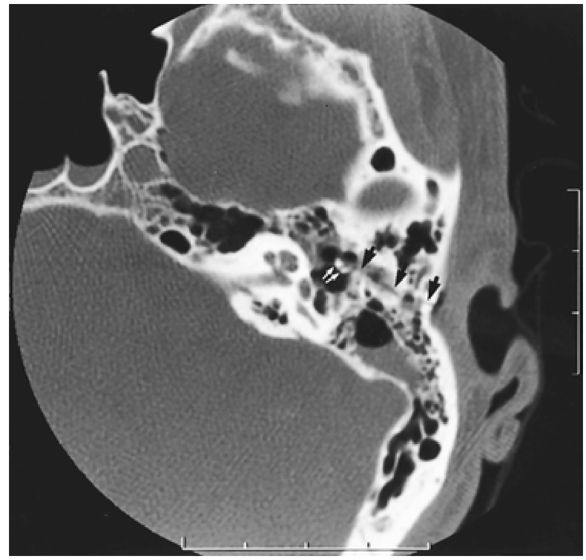


Figure 36–7 *Longitudinal temporal bone fracture, axial computed tomography. A fracture line traverses the temporal bone, running parallel to the long axis of the bone (black arrows). The facial nerve was uninvolved, but the malleoincudal articulation was dislocated (not shown). However, note the abnormal configuration of the ossicular chain in the mesotympanum (white arrows).*

Pneumolabyrinth (air in the membranous labyrinth) generally indicates fracture of the oval window. Fluid in the middle ear may, in the appropriate clinical setting, indicate the presence of a perilymphatic fistula. Penetrating EAC injuries (such as Q-Tip injuries) can result in

Figure 36–8 *Transverse temporal bone fracture, axial computed tomography. A fracture line traverses the temporal bone, running perpendicular to the long axis of the bone (black arrows). The fracture traverses the vestibule (v) and extends into the geniculate ganglion (arrowhead).*

protrusion of the stapes footplate and crura into the vestibule.

VASCULAR LESIONS

Evaluation of vascular masses behind an intact TM invokes a specific differential diagnosis that can often be resolved with appropriate imaging studies. When a middle ear lesion is found on otoscopy, the evaluation should begin with CT. Most of the time, the noncontrast examination will differentiate between an aberrant carotid canal or carotid artery aneurysm, large dehiscent jugular bulb with or without a diverticulum, and a glomus tympanicum tumor. If the noncontrast study is unclear, intravenous contrast injection may be useful to confirm or exclude carotid or jugular origin of the lesion by demonstrating enhancement equivalent to that of blood pool. A glomus tumor will enhance less than blood pool on immediate postcontrast images, and more than inflammatory disease. MRI/MRA may be useful, but they are subject to the limitations mentioned earlier. Location of a discrete mass along the medial wall of the middle ear cavity suggests a glomus tympanicum. A large glomus tympanicum bulges laterally against the TM and is associated with a surprising lack of ossicular erosion for its size (**Fig. 36–9**). Integrity of the osseous skull base is critical in determining extratympanic spread of a glomus tympanicum and in making the diagnosis of glomus jugulare. An irregular lytic, permeative, and often subtle destructive pattern of bone involvement will be seen. If bone involvement is found in a patient with glomus tympanicum and early jugular bulb invasion has occurred, the MRI will be negative. In this situation, a postcontrast CT scan will demonstrate a filling defect in the jugular bulb representing the tumor nodule.

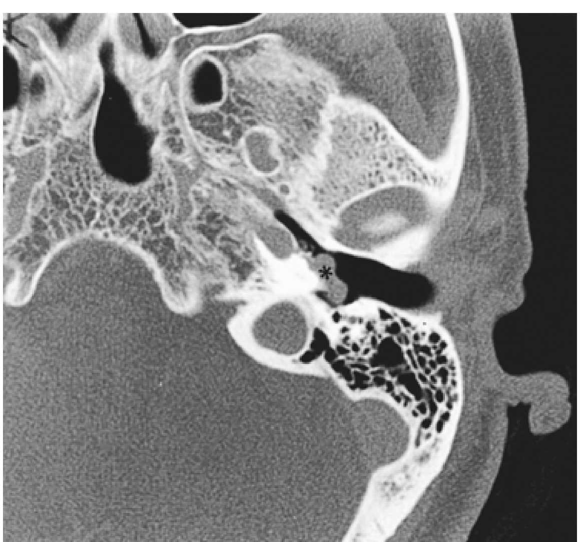


Figure 36–9 *Glomus tympanicum, axial computed tomography. A bilobed, smooth, well-circumscribed mass is seen along the medial wall of the middle ear, in the hypotympanum (asterisk). The adjacent bone is intact.*

Subjective tinnitus with a normal middle ear exam poses a different clinical problem. An MRI is the imaging modality of choice to evaluate the posterior fossa and skull base to exclude neoplasm and vascular lesions such as aneurysm, arteriovenous malformation (AVM), and dural arteriovenous fistula (AVF). If the MRI is completely normal, the MRA will invariably be negative as well. Dural arteriovenous fistulas can go undetected on all cross-sectional imaging studies including MRA, and conventional angiography is necessary to confirm this diagnosis. However, without venous hypertension or parenchymal varices, both of which would be detected on MRI, a dural AVF is unlikely to bleed. Glomus jugulare tumors large enough to produce tinnitus will be well delineated on MRI (**Fig. 36–10**). These chemodectomas have a characteristic hyperintense, speckled appearance on T2W MRI and demonstrate marked contrast enhancement. MRA will delineate many but not all of the arterial feeders. If surgery is selected for treatment, catheter angiography will be necessary for vascular mapping and preoperative embolization.

PETROUS APEX

Lesions of the petrous apex raise yet another differential diagnosis. Both CT and MRI are usually indicated for complete evaluation: the CT defines the character of bony involvement, and the MRI defines the character of the lesion contents.

Petrous apicitis (including Gradenigo's syndrome) is characterized by the presence of a pneumatized but opacified petrous apex, with identification of preserved septations on CT; as the infection progresses, cortical and septal destruction can occur, representing osteomyelitis. High signal intensity will be noted on the T2W MRI with enhancement on the postcontrast T1W images. Outlet obstruction of petrous air cells can result in a petrous apex mucocele. This is a smooth expansile lesion with a thin bony rim. A mucocele can demonstrate either low or high signal intensity on T1W MRI and is generally high in signal intensity on T2W images. The pattern of signal intensity reflects the degree of hydration and protein content of the mucocele. Cholesterol granulomas (**Fig. 36–11**) may occur in the petrous apex and can be identical to mucoceles on CT.

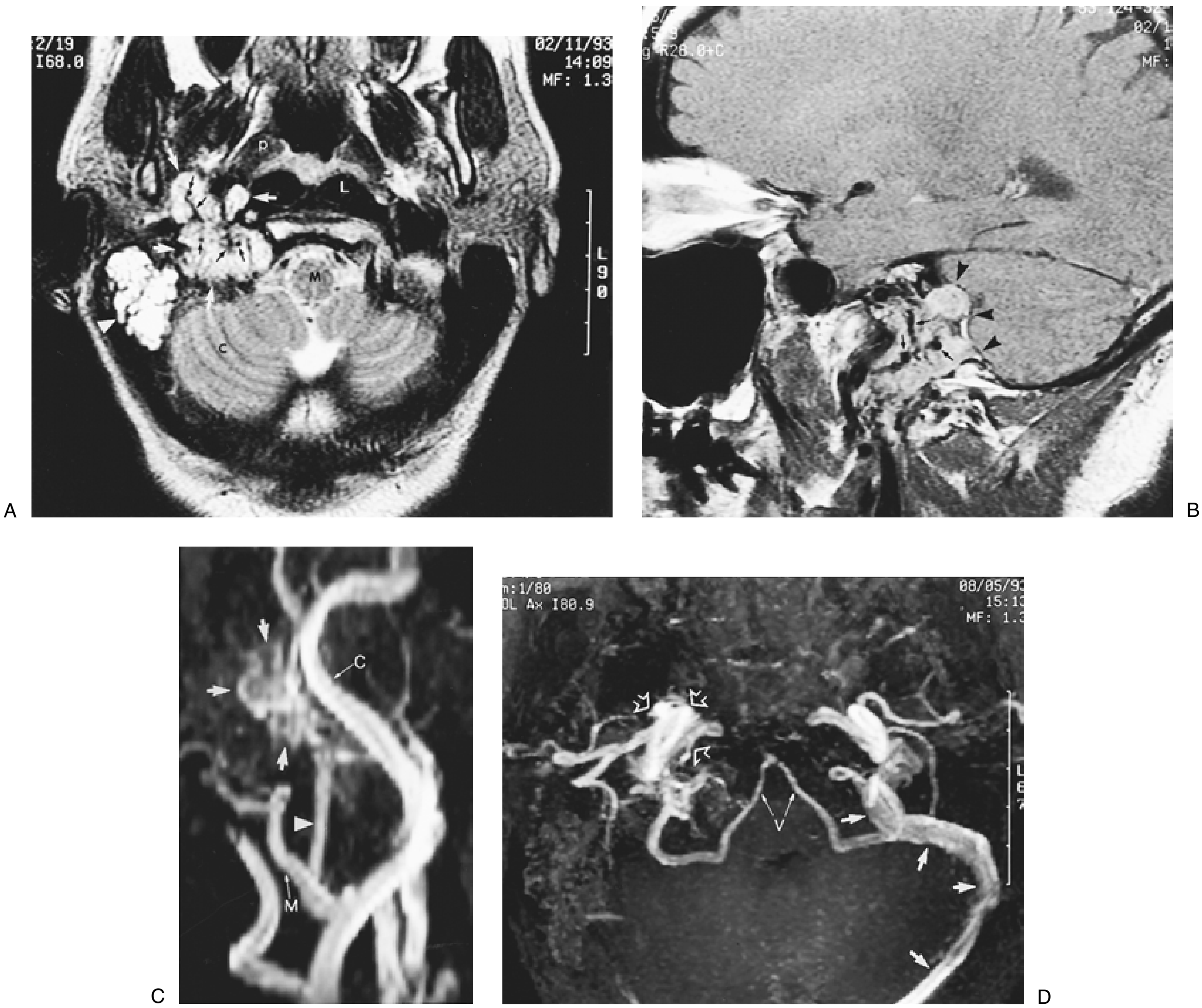


Figure 36–10 Glomus jugulare. ***(A)*** Axial T2-weighted magnetic resonance imaging (MRI). A lobulated hyperintense mass is seen in the right skull base, centered at the jugular bulb (white arrows). Numerous rounded and linear low-signal foci represent flowing blood in the feeding vessels (black arrows). Anterior and inferior extension has obstructed the eustachian tube, and high-signal fluid has accumulated in the mastoid air cells (white arrowhead). C, cerebellum; M, medulla; L, longus colli muscle; p, torus tubarius and palatine musculature. ***(B)*** Sagittal T1-weighted MRI with contrast. The lesion enhances densely with gadolinium. Posterior fossa extension is well seen in this projection (arrowheads). Notice that the vessels maintain flow voids (arrows) despite contrast administration. ***(C)*** Magnetic resonance angiography of the neck, time-of-flight technique, oblique projection. The ascending pharyngeal artery is enlarged (arrowhead). A tangle of tumoral vessels is identified (arrows). C, internal carotid artery; M, internal maxillary artery. ***(D)*** Magnetic resonance angiography of the neck and skull base, time-of-flight technique, axial compression image. There is no flow in the right transverse and sigmoid sinuses and jugular bulb due to the glomus tumor. Note the normal venous drainage on the left side (solid arrows). V, vertebral arteries; open arrows, overlap of the cervical carotid arterial system and tumor.

Both lesions are sharply circumscribed and smoothly expansile. Cholesterol granulomas may be larger and more irregular in shape. The characteristic MRI feature of cholesterol granuloma is high signal intensity (homogeneous or heterogeneous) on the T1W sequence due to the presence of hemoglobin degradation products and cholesterol debris. These lesions are hyperintense on T2W MRI and do not enhance after contrast injection. Congenital cholesteatomas (epidermoid cyst) can arise in the petrous apex and are lytic on CT, with sharply defined margins. On MRI, these lesions are low in signal on T1W and high in signal on T2W sequences. They may therefore appear purely cystic or have some degree of increased T1 signal intensity compared with

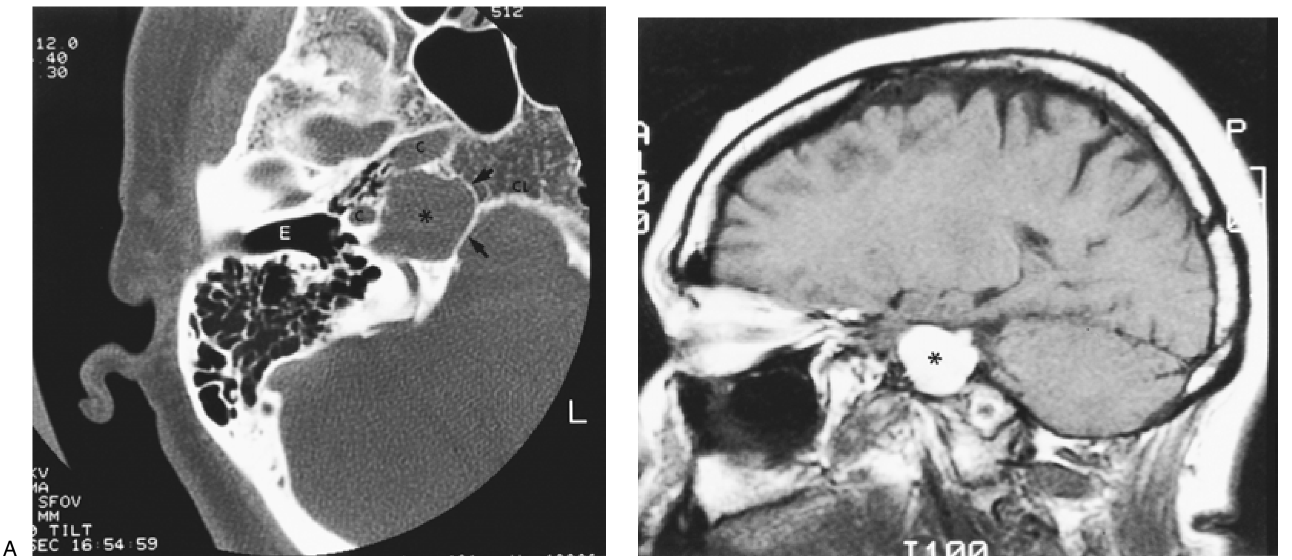


Figure 36–11 *Cholesterol granuloma.* ***(A)*** *Axial computed tomography (CT).* ***(B)*** *Sagittal T1-weighted magnetic resonance imaging (MRI). There is a well-circumscribed rounded lesion in the right petrous apex (asterisk). The thin but intact cortical margin on CT (arrows) indicates the slow expansile character of the lesion that allows the bone to remodel as it enlarges. This type of expansion can displace and distort the carotid canal without eroding into it. High-signal intensity on the noncontrast T1-weighted MRI represents hemorrhagic degradation products and cholesterol crystals, and is the most characteristic imaging feature of this lesion. C, petrous carotid canal; E, external auditory canal.*

CSF. Aneurysms of the petrous carotid artery are rare. Fusiform dilation of the carotid canal will be present. If patent, flow characteristics will be visible on MRA; if thrombosed, clot will be identified on MRI, with stenosis or occlusion on MRA. Primary and metastatic bone lesions may also involve the petrous apex. These will be seen as irregular destructive lesions with ill-defined margins on CT. MRI will confirm the cellular nature of the lesion: hypointense to isointense on T1W and isointense to hyperintense on T2W sequences, with varying degrees of contrast enhancement.

CEREBELLOPONTINE ANGLE

Acoustic neuroma (vestibular schwannoma) is the most common lesion of the CPA. These lesions may be entirely intracanalicular or both intracanalicular and cisternal (**Fig. 36–12**). MRI is the imaging modality of choice to evaluate these tumors, clearly delineating the location and extent of disease. The typical acoustic neuroma is isointense on both T1W and T2W MRI sequences, and enhances densely following contrast administration. When the cisternal component is larger than 2 to 3 cm, intratumoral cystic areas can be seen; arachnoid cysts may be found at the periphery of large acoustic neuromas. In patients who are unable to undergo MRI, CT of the posterior fossa with contrast injection can be performed. Small intracanalicular

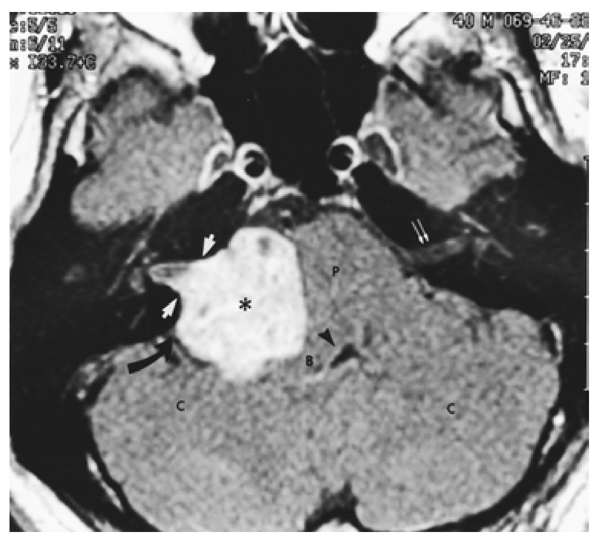


Figure 36–12 *Acoustic neuroma. Axial T1-weighted magnetic resonance imaging following intravenous contrast injection. There is a bulky, enhancing mass lesion (asterisk) in the right cerebellopontine angle (CPA), internal auditory canal (IAC), and a widened porus acusticus (arrows). It extends ~1 cm into the IAC; the 3 cm cisternal component has compressed the pons (P), brachium pontis or middle cerebellar peduncle (B), and cerebellar hemisphere (C) with partial effacement and distortion of the fourth ventricle (arrowhead). The lesion makes an acute angle with the petrous bone (curved arrow). Note the normal IAC and porus on the left side (double arrows).*

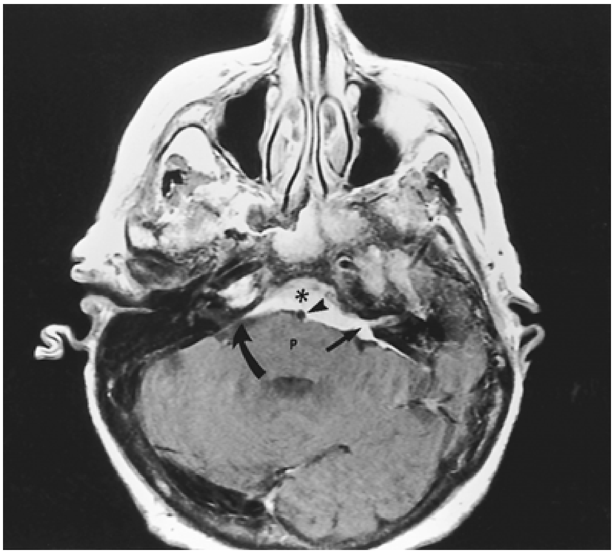


Figure 36–13 *Meningioma. Axial T1-weighted magnetic resonance imaging following intravenous contrast injection. There is a broad, flat, enhancing lesion in the posterior fossa centered at the clivus and posterior petrous bone (asterisk). The mass occupies the prepontine cistern, compressing the pons (P), and partially encasing the basilar artery (arrowhead). It extends into the left cerebellopontine angle and internal auditory canal (arrow) but does not widen the porus acusticus. The lesion makes an obtuse angle with the petrous bone (curved arrow).*

tumors will be missed; bony erosion, however, will be readily detected. A cisternal component larger than 3 to 4 mm will be seen on a high-quality CT performed with contrast administration. Far less common than acoustic neuroma, but the second most common CPA mass, is the meningioma (**Fig. 36–13**). These tumors are benign, arise from the leptomeninges, and will therefore be broad based along the posterior wall of the petrous bone. A rim of dural enhancement frequently surrounds a meningioma and is called the "dural tail" sign. This finding is suggestive of meningioma, but it can be seen in association with other lesions as well. At the margin of a meningioma, an obtuse angle is characteristically formed between tumor and petrous bone, whereas an acoustic neuroma makes an acute angle as it balloons out of the IAC. Meningiomas can extend into the IAC and mimic schwannoma.

Arachnoid and epidermoid cysts are occasionally found in the CPA. Arachnoid cysts form within the leaves of the arachnoid membrane and are similar in density and signal intensity to CSF, and do not enhance with contrast. An epidermoid cyst can be similar in appearance to an arachnoid cyst. Epidermoid cysts exhibit high signal on the diffusion-weighted sequence, while arachnoid cysts are low in signal. The epidermoid cyst grows along paths of least resistance, insinuating around and between various structures in the subarachnoid space. An arachnoid cyst will displace structures as it enlarges.

CONGENITAL MALFORMATIONS

A complete discussion of congenital anomalies is well beyond the scope of this chapter. Here we will consider two abnormalities most frequently encountered: microtia with aural atresia, and the Mondini's malformation.

Microtia is frequently associated with external canal stenosis or atresia as well as a spectrum of anomalies involving the ossicular chain and facial nerve canal. A CT examination of the temporal bone is indicated if and when surgical correction of the conductive hearing loss is planned. The thickness of the atresia plate, degree of EAC stenosis, and fibrous content of a stenotic canal will be clearly delineated. The middle ear cavity will often be narrow, and there is usually some abnormality of the ossicular chain. Most commonly seen is clumping of the malleus and incus into an amorphous mass that is then fused to the atresia plate (**Figs. 36–4** and **36–14**). The course of the posterior tympanic facial nerve canal, the posterior genu, and

Figure 36–14 *Microtia and aural atresia, coronal computed tomography posterior to* **Fig 36–9**. *The short incudal process is seen on this posterior slice. The facial nerve is aberrant with the descending segment (open arrow), located at the anterior rim of the lateral semicircular canal (L) rather than the tympanic segment. The crista falciformis is well delineated in A (black arrow). C, basal turn cochlea; S, superior semicircular canal.*

Figure 36–15 *Mondini malformation. Axial computed tomography. The cochlea (c) is dysplastic. There is no septation between the apical and middle turns, so that they appear as a single sac. The basal turn was completely normal on an inferior slice (not shown). The vestibule is slightly large (V), and the vestibular aqueduct is dilated (arrows). The internal auditory canal (asterisk), semicircular canals, and middle ear structures are normal.*

the descending canal is frequently abnormal, with a foreshortened tympanic segment, anterior location of the genu, and a short, anteriorly angulated mastoid segment.

The Mondini's malformation is a common anomaly of the cochlea, due to an interruption of the developmental sequence at ~7 weeks of gestation. At this point, the cochlea has elongated to produce 1½ turns: the basal turn is fully formed, but the separation between the middle and apical turns has not yet occurred. Therefore, the Mondini's malformation is characterized by incomplete septation of the middle and apical cochlear turns. The CT scan confirms the presence of a normal basal turn and round window. The middle and apical turns are seen as a single sac without the intervening osseous spiral lamina (**Fig. 36–15**). The original description of this entity by Mondini included dilation of the vestibular aqueduct; in practice, this anomaly may or not be associated.

SUMMARY

Interpretation of temporal bone imaging studies is complex, requiring the integration and triangulation of many facts and findings. The key to understanding temporal bone imaging lies in realizing the strengths and limitations of each imaging modality; that is, understanding what structures may or may not be delineated, and what questions may or may not be answered. Beginning a radiographic evaluation with a definite or proposed clinical diagnosis is extremely useful and can be used to direct the initial examination (**Table 36–1**). Using this method, 80 to 90% of single modality studies will be diagnostic, necessitating a second study in the minority of cases.

Table 36–1 Comparison of Computed Tomography and Magnetic Resonance Imaging for Temporal Bone Imaging

Clinical Diagnosis or Condition	CT	CT+ Contrast	MRI
Inflammatory disease–middle ear otitis, cholesteatoma, fistula intracranial complications	**Yes***	No	No
Trauma	No	Yes	**Yes**
Fracture	**Yes**	No	No
Ossicular dislocation	**Yes**	No	No
Intracranial complications	Yes	No	**Yes**
Vascular lesion–middle ear	**Yes**	Yes	Yes
Petrous apex	**Yes**	No	**Yes**
CPA	No	Yes	**Yes**
Congenital anomalies	**Yes**	No	No

*Bold lettering indicates the preferred modality.
CPA, cerebellopontine angle; CT, computed tomography; MRI, magnetic resonance imaging.

Suggested Readings

Jackler RK, Luxford WM, House WF. Congenital malformations of the inner ear: a classification based on embryogenesis. Laryngoscope 1987;97(suppl 40):2–24

Mayer TE, Brueckmann H, Siegert R, Witt A, Weerda H. High-resolution CT of the temporal bone in dysplasia of the auricle and external auditory canal. AJNR Am J Neuroradiol 1997;18:53–65

Osborn AG. Brain tumors and tumor like masses: cerebellopontine angle masses, internal auditory canal and temporal bone masses. In: Osborn AG, ed. Diagnostic Neuroradiology. St. Louis: Mosby Yearbook; 1994:437–450

Swartz JD, Harnsberger HR. Imaging of the Temporal Bone. 3rd ed. New York: Thieme Medical Publishers; 1998

Valvassori GE, Buckingham RA. Radiology of the temporal bone. In: Valvassori GE, Buckingham RA, Carter BL, Hanafee WN, Mafee MF, eds. Head and Neck Imaging. New York: Thieme Medical Publishers;1988:1–8

Weissman JL. Hearing loss. Radiology 1996;199:593–611

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 716.

1. For evaluation of tinnitus, the optimal imaging modality is
   A. Computed tomographic (CT) scan with intravenous (IV) contrast
   B. Positron emission tomographic (PET) scanning
   C. Angiography
   D. T2-weighted magnetic resonance imaging (MRI)
   E. MRI with IV contrast
2. Petrous apex lesions include cholesterol granulomas, cholesteatomas, and mucoceles. These lesions are optimally evaluated with
   A. T2-weighted MRI
   B. MRI with IV contrast
   C. CT scan
   D. Combination of plain CT and MRI with IV contrast
3. Temporal bone trauma causing facial nerve paralysis is best evaluated by
   A. Surgical exploration
   B. CT scan with IV contrast
   C. MRI
   D. CT scan without contrast high resolution

This page intentionally left blank

Part III

THE NOSE, OLFACTION, AND THE SINUSES

Chapter 37

DEVELOPMENT OF THE NOSE

BRADLEY J. GOLDSTEIN AND THOMAS R. VAN DE WATER

Nasal development is intimately associated with the development of the palate and midface. Development of these structures is a complex process involving interaction of the nasal placode, frontonasal prominence, and branchial arch tissue. An understanding of the embryology of this region will allow the surgeon to better appreciate both normal anatomy and congenital pathology.

NORMAL DEVELOPMENT

By the fourth embryonic week the human embryo is ~5 mm in length, and the early face consists of the primitive oral opening, termed the stomodeum, the frontonasal prominence, the paired nasal placodes, and the maxillary and mandibular prominences of the first branchial arch (**Fig. 37–1**). Development of the face from these primordia occurs mainly between weeks 5 through 8. The frontonasal prominence arises from the mesenchymal tissue ventral to the developing telencephalon and forms the cranial boundary of the stomodeum. The caudal boundary of the stomodeum is formed by the mandibular prominences of the first branchial arch and will not be discussed here. The lateral boundaries are formed by the paired maxillary prominences of the first branchial arch. Mechanistically, important developmental events involve the induction of embryonic tissue by nearby structures; the molecular bases for these processes are beyond the scope of this chapter. For simplicity, the structures derived from the above-mentioned primordia will be considered separately, although their development is closely associated.

FRONTONASAL PROMINENCE AND FIRST BRANCHIAL ARCH

The frontonasal prominence will give rise to both external and internal structures. Externally, the forehead, dorsum of the nose, and apex of the nose will arise from the frontonasal prominence. Inferiorly, the frontonasal prominence will divide into the lateral and the medial nasal prominences as a consequence of the presence of the paired nasal placodes (**Fig. 37–1**). The nasal placodes arise as thickenings of embryonic ectoderm, invaginate to form the nasal pits, and will be considered in detail here. The tissue lateral to the nasal placode is termed the lateral nasal prominence, and it is separated laterally and inferiorly from the first branchial arch by a furrow termed the nasolacrimal groove. The lateral nasal prominence will give rise externally to the alae, or sides of the nose. The paired

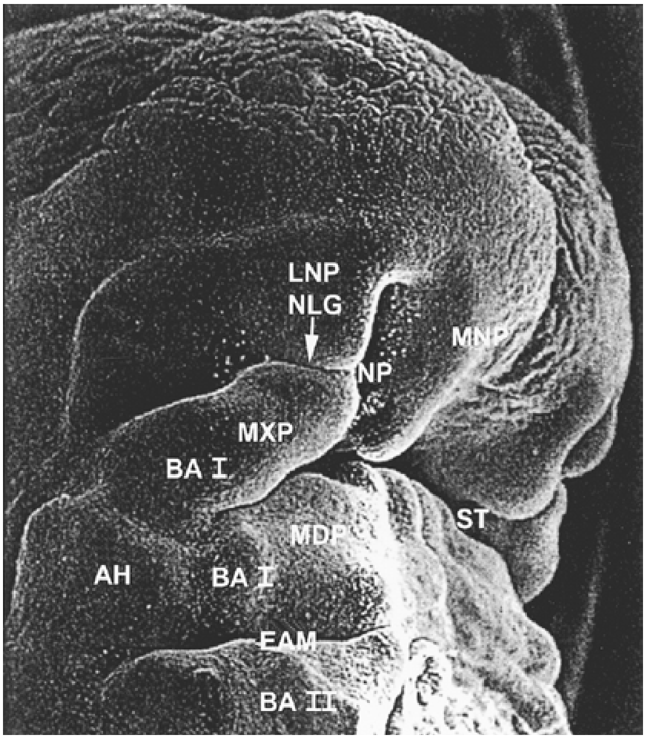


Figure 37–1 *An oblique view of the craniofacial region (scanning electron micrograph) of a stage 16 (approximately gestation day 41) human embryo. The facial structures that contribute to the formation of the nose are present: the medial nasal process (MNP); lateral nasal process (LNP); nasal placode (NP); maxillary process (MXP); nasal lacrimal groove (NLG). Other craniofacial structures present are: first branchial arch (BA I); second branchial arch (BA II); stomodeum (ST); mandibular process (MDP); auricular hillock (AH); and groove of the external auditory meatus (EAM).*

medial nasal prominences will grow medially and inferiorly and will merge with each other as well as with the maxillary prominences of the paired first branchial arches. This tissue is then termed the intermaxillary segment. Externally, the philtrum of the lip will arise from the intermaxillary segment. Internally, the intermaxillary segment forms a structure termed the primary palate. The premaxilla, containing the incisors, arises from the primary palate. The primary palate will give rise to only the very small portion of the hard palate that is anterior to the incisive foramen. The remainder of the palate will be derived from first branchial arch tissue, termed the secondary palate, and will be discussed in detail later. Interiorly, the medial nasal prominences give rise to the nasal septum. The septum develops as a down-growth from the fused medial nasal prominences. The septum will eventually join the merging lateral palatine processes of the first branchial arches to separate definitively the nasal fossae from the oral cavity.

The vomeronasal organ of Jacobson exists transiently during development in the anteroventral portion of the nasal septum. Just above the primitive secondary palate a pair of diverticulae invaginate and become subsequently lined by cartilage and a sensory neuroepithelium. The neuroepithelium contains bipolar sensory neurons morphologically similar to olfactory receptor neurons of the developing first cranial nerve. These vomeronasal neurons project their axonal fibers as a vomeronasal nerve to a region of the developing telencephalon just caudal to the developing olfactory bulbs, termed the accessory olfactory bulb. The vomeronasal organ is maximally developed by ~6 months and then regresses, usually completely. The only adult remnants are normally the vomeronasal cartilages. Of note, the vomeronasal organ is well developed and functional in other mammals. Murine vomeronasal neurons have been shown to express putative pheromone receptor genes, and the vomeronasal organ is an important sensory system involved in reproductive behaviors in many species.

The first branchial arch: the first branchial arch forms the paired maxillary prominences. Externally, the maxillary prominences will give rise to the lateral portions of the upper lip. It is important to note that the tissue of the primitive face and cheeks, although derived from the maxillary prominence of the first branchial arch, will be invaded by mesenchyme of the second branchial arch, which will give rise to facial musculature. Recalling that the nerve of the second branchial arch is cranial nerve VII (CN VII), it is clear that the muscles of facial expression are then supplied by the seventh nerve. Muscles of mastication, however, arise from first branchial arch mesenchyme and are therefore supplied by the nerve of the first branchial arch, i.e., CN V. Internally, the maxillary prominences will give rise to the hard and soft palate, posterior to the incisive foramen, as already mentioned. Palate development occurs from late in the fifth week until approximately the 12th week. The primary palate, derived from the intermaxillary segment of the frontonasal process, discussed earlier, gives rise to the anterior extent of the hard palate. The remaining hard and soft palate derive from the first branchial arch tissue priordium, termed the secondary palate. The lateral palatine processes, or shelves, are paired horizontal extensions of mesoderm that form from the maxillary prominences internally. The lateral palatine processes extend medially and eventually fuse in the midline, where the developing nasal septum also fuses. The process of fusion occurs in an anterior to posterior fashion, starting at the primary palate and ending at the uvula. Fusion occurs from about the seventh to 12th week. The portions of the lateral palatine processes that fuse beyond the nasal septum do not undergo infiltration of osseous tissue from the maxillae and palatine bones, and become the soft palate and uvula. Muscles of the soft palate are derived from myotome cells that first invade the fourth arch and subsequently migrate to the palate, carrying their innervation from CN X. The pathology of clefts will be discussed later; however, it is apparent from the discussion of normal development that the pathogenesis of cleft lip and cleft palate comprises two distinct developmental entities.

The nasolacrimal groove, separating the maxillary prominence from the lateral nasal prominence, will give rise to the nasolacrimal duct and the lacrimal sac. A thickening of ectoderm, the groove canalizes and ends in the inferior meatus of the lateral nasal wall. The cranial end of the developing duct establishes connections with the conjunctival sac, and the lacrimal gland forms as an ectodermal proliferation at the superior lateral portion of the orbit, superficial to the orbital septum. An atresia of the caudal portion of the nasolacrimal duct may arise from incomplete canalization.

Nasal Placode

The nasal placodes are paired ectodermal thickenings above the stomodeum present by the fourth week. Interaction among cells of the frontonasal prominence, the first branchial arch, and the nasal placode result in the formation of nasal pits and subsequent nasal sacs that give rise to the nasal cavities, septum, and lateral nasal walls. Specialized cells of the nasal placode give rise to the olfactory neuroepithelium that will line the posterodorsal regions of the nasal fossae. As the nasal sac invaginates, an oronasal septum forms and transiently separates the nasal cavity from the oral cavity. The oronasal septum then ruptures at about week 8, forming the primitive choanae. Just caudal to the oronasal septum is the origin of Rathke's pouch, which will give rise to the anterior pituitary. As the tissue of the frontonasal prominence and maxillary prominence continues to develop around the nasal sacs, the superior, middle, and inferior conchae are formed on the lateral nasal walls. Chondrification of a condensation of mesenchyme surrounding the developing nasal cavity forms the nasal capsule. At approximately week 7, the inferior and middle conchae are formed from anlagen, termed the maxilloturbinal and first ethmoturbinal, respectively. The uncinate process is also formed at this time. During the eighth week, the superior concha is formed from the second and third ethmoturbinals. The mucosa lining the nasal cavities will form the upper airway and develops mucus-secreting ciliated columnar cells. The exception is the placodally derived olfactory neuroepithelium. An epithelial plug occupies the external nares from approximately week 8 through week 24, after which time it is normally resorbed. Stenosis or atresia of the external nares can result from incomplete resorption.

Even before the nasal placode invaginates, cells of the placode begin to express genes involved in neuronal development. The earliest events involve expression of *Hes*1/*Hes*5 genes, which appear to regulate expression of the proneural gene *Mash*1 in olfactory neuronal progenitor cells derived from the nasal placode. Placodally derived cells lining the superior turbinate, the posterodorsal middle turbinate, and the posterodorsal region of the nasal septum develop into the olfactory neuroepithelium. This epithelium houses the cell bodies of the olfactory receptor neurons of CN I. The axonal projection of these bipolar olfactory receptor neurons comprises CN I and traverses the cribriform plate of the ethmoid bone to synapse with mitral and tufted cell dendrites in the glomeruli of the olfactory bulbs of the brain. The olfactory bulbs exist as telencephalic outgrowths. Innervation by axons from olfactory receptor neurons of the nose appears to be necessary for complete development of the olfactory bulbs. Also, specialized cells from the nasal placodes express markers for gonadotropin-releasing hormone (Gn-RH) cells and migrate early in development from the nasal placodes to the region of the hypothalamus. The development and migration of these cells appears to be a necessary early component in normal hypothalamic development. Understanding this relationship provides insight into the etiology of hypogonadotropic hypogonadism with anosmia (Kallmann's syndrome). Developmental anomalies related to the nasal placode will be discussed later. The olfactory neuroepithelium of the nose retains embryonic characteristics postnatally. The production of new olfactory receptor neurons from mitotic progenitor basal cells continues through adulthood, replacing senescent receptor neurons. This robust capacity for ongoing neurogenesis contrasts sharply with other regions of the adult nervous system. The neurobiology of the olfactory system is described in detail elsewhere in this text (see chapter 40).

Paranasal Sinus Development

Paranasal sinus development begins late in development and continues postnatally. From the lateral nasal wall, 3 to 4 mm maxillary sinuses are present at birth, with ostia into the middle meatus. Also, a small number of anterior and posterior ethmoidal air cells are present. At around 2 years of age, frontal sinuses are formed from the growth of the most anterior ethmoids into the frontal bone. At about the same time, the sphenoid sinuses are formed from the growth of the most posterior ethmoids into the sphenoid bone. Paranasal sinus growth continues into puberty, and maxillary sinuses reach full development only after all permanent dentition has erupted.

DEVELOPMENTAL ANOMALIES

Anomalous development involving the nose, face, and central nervous system (CNS), in isolation or in combination, or as elements of certain syndromic conditions, is of clinical importance to the otolaryngologist. Problems include congenital nasal masses, clefts, atresias, and olfactory sensory dysfunction.

Choanal Atresia

Choanal atresia occurs when the oronasal membrane fails to rupture during the sixth week of development. The defect may be unilateral or bilateral and may be membranous or, more commonly, bony. Because newborns are obligate nose breathers, a bilateral choanal atresia presents as immediate respiratory distress and is usually managed with intubation. Some forms of choanal atresia have been reported to be transmitted in an autosomal recessive manner. Definitive surgical repair may be accomplished via a transnasal or transpalatal approach, depending on the specific defect. Choanal atresia may be associated with the CHARGE syndrome (*c*oloboma of the eye, *h*eart anomaly, choanal *a*tresia, *r*etardation, and *g*enital and *e*ar anomalies), associated with a micodeletion in chromosome 22q11. Because of airway instability, these patients usually undergo neonatal tracheotomy, with delayed definitive repair.

Other Anomalies Related to Early Defects

Rare anomalies related to very early defects in nasal placode formation have been reported. Complete failure of placode formation results in absence of the nose; presence of a single nostril may result from formation of only one nasal placode. A bifid nose results from incomplete merger of the medial nasal prominences. Various facial clefts related to nasal development may occur, usually in association with gross craniofacial malformation, and are extremely rare. For instance, failure of frontonasal prominence mesenchyme to merge with elements of the maxillary prominence results in an oblique facial clefts from the upper lip to the medial orbit, with persistent nasolacrimal grooves.

Kallmann's Syndrome

A failure of certain neurons to migrate from the nasal placode to the developing hypothalamus results in Kallmann's syndrome. There are both X-linked and autosomal forms of this disease. Males with X-linked disease have anosmia secondary to failure of olfactory bulb development, hypogonadotropic hypogonadism, micropenis, cryptorchidism, unilateral renal agenesis, and other defects. The *KAL* gene located on the X chromosome at Xp22.3 appears to encode a cell adhesion protein required for proper neuronal migration. Hormonal therapy is available for treatment of the hypogonadism and infertility.

Congenital Nasal Masses

Congenital nasal masses include cysts, neoplasms, meningoceles, encephaloceles, and vascular lesions. These anomalies may be associated with other craniofacial or systemic deformities. Evaluation of midline nasal masses should involve either or both computed tomographic and magnetic resonance imaging scans prior to surgical treatment to determine intracranial extension and contents of the mass [i.e., brain, cerebrospinal fluid (CSF), and vascular structures].

A dermoid cyst is a congenital benign neoplasm, derived from ectodermal components, containing keratinizing squamous epithelium and adnexal skin structures such as sebaceous glands and hair follicles. Dermoid cysts are thought to arise from a failure of obliteration of dural projections through the fonticulus frontalis, the space that exists between the frontal bones prior to their fusion. The cyst often presents at birth on the nasal bridge with a slight dimple and sinus tract; however, it may be located anywhere from the columella to the nasion. Dermoids may involve only the skin and nasal bones, or they may have a true dural connection.

Meningoceles or encephaloceles may also present as masses on the nasal bridge or may be located in the superior nasal cavity. These sacs communicate with the ventricles and contain CSF. The meningocele contains meninges, and the encephalocele also contains normal brain tissue. Vascular structures may be present within the herniation.

Central nervous system heterotopia may present on the nasal bridge or submucosally in the nasal cavity. These congenital rests of CNS tissue are of varying growth rate. Histologically, they usually contain glial elements in a fibrillar stroma; therefore, they are often designated as gliomas, although this entity is not a true glial neoplasm, such as an astrocytoma or oligodendroglioma. Gliomas, meningoceles, encephaloceles, and dermoids exist in the differential diagnosis of the congenital midline nasal mass. Furstenburg's test, in which the jugular vein is compressed to increase CSF pressure, may produce enlargement of meningoceles and encephaloceles.

Teratomas are true neoplasms composed of multiple cell types, arising from totipotent cells. Half of all head and neck teratomas occur in the nose. The teratoma may present as a polypoid mass in the nasopharynx. Teratomas may be associated with other congenital anomalies.

The differential diagnosis of nasopharyngeal masses also includes cysts of Rathke's pouch or infection of Tornwaldt's bursa. Rathke's pouch is a remnant from the ectodermal invagination during anterior pituitary development. Tornwaldt's bursa represents notochord remnant inferior to Rathke's pouch.

Cleft Lip and Cleft Palate

Cleft lip and cleft palate etiology is related to nasal development, as described earlier in this chapter. Cleft lip, with or without cleft palate, affects ~1 in 1000 live births, with ethnic variation. Cleft lip and cleft palate are the most common malformations in the head and neck region. Because of an association with many malformation syndromes, the physician must perform a thorough assessment for other congenital anomalies. Autosomal dominant syndromes associated with cleft lip and/or palate include Apert's, Stickler's, Treacher Collins, and Waardenburg's syndromes. The cleft population overall is distributed as follows: clefts of the lips, alveolus, and palate, 45%; clefts of the secondary palate only, 30%; clefts of the alveolus and lip, or lip only, 25%.

Several classification systems for clefts have been described. It is useful to describe the defects as primary versus secondary, complete versus incomplete, and unilateral versus bilateral, and to document this diagrammatically. Primary clefts involve tissue derived from the embryonic primary palate (the hard palate anterior to the incisive foramen and the maxillary alveolus and lip), and secondary clefts involve tissues derived from the secondary palate (the uvula, soft palate, and hard palate posterior to the incisive foramen). Complete clefts are those having extension to the nasal floor; incomplete clefts range from muscular diastasis with intact epidermis or mucosa to the presence of only a thin tissue band bridging the medial and lateral structures. Although the precise mechanisms causing cleft lip and palate are incompletely understood, the basic problem involves defects or interference in tissue fusion early in development. These problems are thought to have multiple etiologies, both genetic and environmental in nature. The primary palate forms during weeks 4 through 7, as already described, and the secondary palate forms later, during weeks 7 through 12, with the palatal shelves fusing in an anterior to posterior fashion ending at the uvula. Failure in the merger of the medial nasal process with the maxillary process is thought to underlie cleft lip formation. Failure of the lateral palatine processes to meet and fuse is thought to underlie cleft palate formation.

Problems facing the cleft lip or cleft palate patient vary with the severity of the defect. Issues to be considered include feeding, speech development, velopharyngeal insufficiency, midface development, otitis media secondary to eustachian tube dysfunction, and cosmesis. Broadly, the goals of cleft lip therapy are acceptable cosmesis and facilitation of closure of the alveolar arch as needed. Goals of cleft palate therapy involve closure of the defect without interfering with maxillary arch growth, facilitation of velopharyngeal closure, and speech development. A discussion of the techniques for surgical correction is beyond the scope of this chapter; see Suggested Readings for this information.

SUMMARY

Ectodermal thickenings termed the nasal placodes invaginate and, in combination with the first branchial arch and frontonasal prominence tissue, give rise to the structures of the nose and palate. In addition to the structural components of the septum and lateral nasal wall, this development includes the formation of the peripheral olfactory sensory system. The cell bodies of neurons of the first cranial nerve are situated in a neuroepithelium in the posterodorsal portion of the nasal fossae. Paranasal sinus development begins late in development and persists postnatally. Relatively common clinical entities occur secondary to malformations in nasal development. These include choanal atresia, congenital nasal masses, and congenital clefts. Cleft lip and palate are the most common head and neck malformations.

Suggested Readings

Bailey BJ. Head and Neck Surgery: Otolaryngology. 2nd ed. New York: Lippincott-Raven; 1998:1177–1186, 1203–1220

Cummings CW, et al. Otolaryngology Head and Neck Surgery. Vol 5. 3rd ed. St. Louis: Mosby; 1998;92–103, 133–144

Moore K. The Developing Human. 3rd ed. Philadelphia: Saunders; 1982:197–215

Sweeny LJ. Basic Concepts in Embryology. New York: McGraw-Hill; 1998:214–220

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. The ectodermal thickening that will give rise to olfactory receptor neurons is the
   A. Nasal placode
   B. Frontonasal prominence
   C. Medial nasal prominence
   D. Rathke's pouch
2. Choanal atresia occurs due to the persistence of
   A. Primary palate
   B. Vomeronasal organ
   C. Nasolacrimal groove
   D. Oronasal septum
3. Fusion of the lateral palatine processes occurs in an anterior to posterior direction
   A. True
   B. False
4. The olfactory neuroepithelium continues to produce new neurons from progenitor cells
   A. Until the fourth week of development
   B. Until the 12th week of development
   C. Until birth
   D. Throughout adulthood
5. A dermoid cyst may contain
   A. Cerebrospinal fluid
   B. Adnexal skin structures
   C. Congenital rests of central nervous system tissue
   D. Teeth

Chapter 38

SURGICAL ANATOMY OF THE NOSE AND PARANASAL SINUSES

Dinesh Mehta and Walter M. Ralph, Jr.

As is true for many organ systems, nasal anatomy is intimately associated with nasal physiology. Despite the many ethnic and gender variations on the theme, the overall framework and structure of the nose facilitate its four basic physiological functions: humidification, respiration, olfaction, and filtration. The architecture of the nose is so involved with its function that alteration of its structure always affects more than one function. With this in mind, it is easy to see how a patient with a nasal structural abnormality will usually present with more than one complaint relating to nasal physiology. Conversely, a conscientious physician must look for related signs and symptoms of altered function when examining the patient with nasal or paranasal complaints.

The anatomy of the nose is grossly divided into external and internal anatomy. Subsequently, within each category are subdivisions based on anatomical designations. These categories are artificial constructs and are designed only to facilitate our understanding of the anatomy of the nose. All nasal components are interrelated and function as a unit to allow us normal nasal function.

AESTHETIC ANATOMY OF THE NOSE

The external appearance of the nose has rhinometric and aesthetic significance. The rhinometric significance is to facilitate the maximum passage of air through the nostrils with minimum resistance. How well it functions in this capacity is influenced by individual as well as ethnic and genetic factors. It can be measured objectively. Conversely, the aesthetic significance of the nose is much harder to quantify because it is not constant but varies with societal norms and personal preferences. With its central location on the face, much emphasis has been placed on its aesthetic significance, and unique descriptive language has developed to aid professionals in communicating about its features and orientation.

SURFACE ANATOMY

For anatomical orientation, the terms *anterior, posterior, superior*, and *inferior* are supplemented by the terms *ventral, dorsal, cephalic*, and *caudal*, respectively. For purposes of illustration, the nose has been described as a "curious creature with its root at the top, its back in front and its wings on the bottom." Several key areas along the midsection of the face and nose have been described and correspond to depressions and elevations, which affect the overall contour of the nose (**Fig. 38–1**). The glabella is the point that marks a smooth prominence on the frontal bone ~1 to 1.5 cm above the radix of the nose. The radix is the uppermost extent of the nose as determined by direct visual inspection. The nasion is a depression just cephalad to the radix that marks the nasofrontal angle. This is not to be confused with the nasofrontal suture line, the suture between the frontal bone and the paired nasal bones. The nasion and the nasofrontal suture line may overlap, but because of underlying subcutaneous fat and muscle, the nasion is usually a few millimeters cephalad of the nasofrontal suture line. The rhinion is the junction between the bony and cartilaginous dorsum. This is often referred to as the "hump" of the nose; depending on ethnic or gender influences, it can be a very prominent landmark on the dorsum. It tends to be a more prominent feature in men than in women.

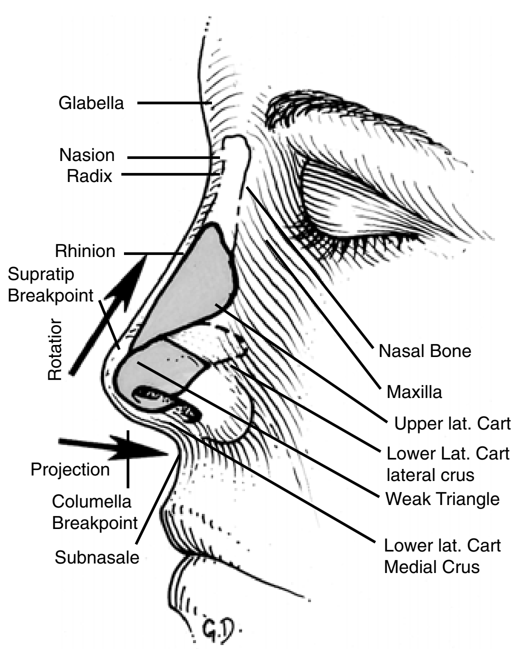


Figure 38–1 *Profile of the nose showing the surface landmarks and the underlying structural components. The black arrows suggest the common terminology used by aesthetic surgeons for resurfacing the nasal profile.*

The nose traditionally has been divided into thirds, the cephalic third extending from the radix to the rhinion, the middle third extending from the rhinion to the weak triangle (a term to be defined later), and the caudal third encompassing everything below. The upper two thirds forms the nasal dorsum and projects caudally from the nasofrontal angle. The nasal bridge consists of the dorsum bounded laterally by the side walls. The caudal third is often referred to as the base or lower third. It is more complicated in its composition and is made up of the lobule, columella, ala, sill, and vestibule.

The lobule is the area just caudal to the weak triangle that gives rise to the tip of the nose. Aesthetically, the "tip" of the nose is not an exact anatomical structure but is defined as the midpoint of a line that connects the two tip-defining points (**Fig. 38–2**). The tip-defining points are anatomical structures formed by the dome segments of the lower lateral cartilages beneath the overlying skin. From the front view, they are seen as reflections of light on the lateral aspects of the lobule. The tip-defining points also form the apex of a region known as the soft triangle, the base and sides of which are formed by the nostril rim and margins of the medial and lateral crus, respectively (**Fig. 38–2**). The soft triangle is formed at the caudal junction of the ala and lobule and is an area where there is no intervening subcutaneous tissue between the dermis of the outer surface and the dermis of the inner vestibular surface of the nostril rim. The cosmetic surgeon applies great surgical significance to the soft triangle. Inadvertent trauma to this area can lead to severe postoperative scarring and deformity.

The presence of the tip-defining points divides the lobule into supratip and infratip segments. The supratip lobule extends from the weak triangle to the tip-defining point. The infratip lobule extends from the tip-defining point to the columella breakpoint (**Fig. 38–1**). As viewed from the side (profile), the point where the plane of the supratip diverges or breaks from the plane of the dorsum is called the supratip breakpoint. This is usually 1 to 2 mm above the tip-defining point. The angle of the supratip breakpoint influences the aesthetic parameter known as tip projection (**Fig. 38–1**).

The ala refers to the fleshy, convex portion of the nose just lateral to either side of the lobule (**Fig. 38–2**). It forms the hood or ceiling of the vestibule, the cavity that forms the entrance to the nasal cavity. The junction of the lobule and ala forms a depression termed the alar-lobular crease. At the lateral aspect of the ala is another depression formed by the junction of the ala and the skin of the face, termed the alar-facial crease. The columella is the midline structure that extends ventrally from the lobule to the area where the nose abuts the upper lip,

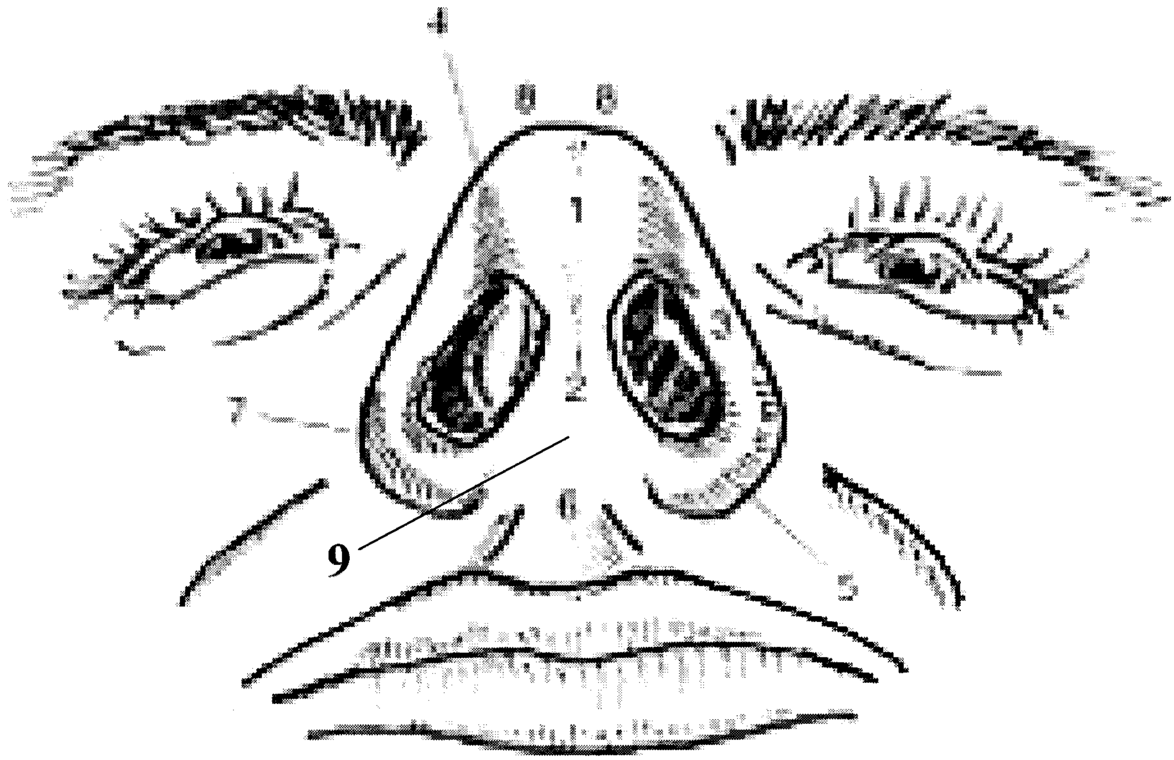


Figure 38–2 *Line drawing of the caudal to cephalic frontal view of nose with important surface landmarks and structures indicated. 1, Infratip lobule; 2, Columella; 3, Ala; 4, Soft triangle; 5, Sill; 6, Philtrum; 7, Alar-facial crease; 8, Tip defining points; 9, Subnasale*

also known as the subnasale. It also forms part of the dividing wall of the vestibule. The sill is the caudal and medial extension of the ala that merges with the subnasale. It forms a shelf of skin that extends from the ala-facial crease to the columella.

STRUCTURAL ANATOMY

The overlying surface anatomy as just described is fashioned as a direct consequence of the underlying framework, which consists of both bony and cartilaginous structures. Forces that alter the form, integrity, or position of segments of this framework directly affect the shape of the surface anatomy of the nose, ultimately affecting physical appearance. Understanding the dynamics of these structural forces is the basis for rhinoplasty procedures. Numerous cosmetic surgical techniques have been developed over the centuries to alter the features of the nose to meet aesthetic norms and personal preferences. A detailed discussion of rhinoplasty is beyond the scope of this chapter. The pyriform aperture is a large, pyramidal-shaped opening in the center of the skeletal midface that forms the bony entrance into the nasal cavity (**Fig. 38–3**). The skeletal frame that gives shape to the aperture is termed the bony vault. It is formed from the paired nasal bones above and the

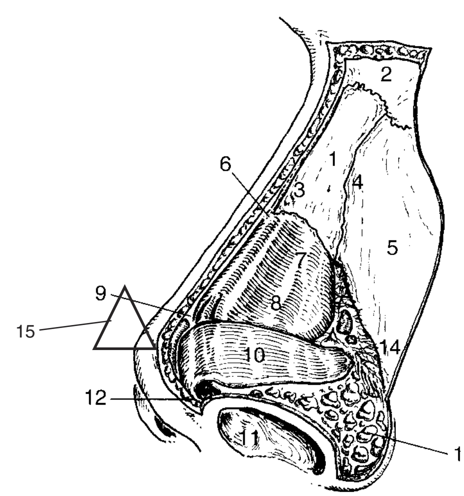


Figure 38–3 *Line drawing of the underlying bony and cartilaginous structures of the dorsum and side walls of the nose. 1, Nasal bone; 2, Frontal bone; 3, Nasal suture line; 4, Nasomaxillary suture line; 5, Maxilla; 6, Rhinion; 7, Upper lateral cartilage; 8, Caudal part of ULC; 9, Septal angle; 10, Lower crus of lower lateral cartilage; 11, Medial Crus LLC; 12, Intermediate crus LLC; 13, Sesamoid cartilage; 14, Pyriform aperture; 15, Weak triangle area*

paired ascending frontal processes of the maxilla laterally (**Fig. 38–3**). The area of articulation of the nasal bones and the frontal bone, termed the nasofrontal angle, forms a gentle curve of ~115 to 130 degrees. The nasolabial angle, formed by the upper lip and columella, likewise defines the projection of the tip of the nose (**Fig. 38–1**). These angles play an important role in cosmetic and reconstructive rhinoplasty, in that a more acute angle is associated with masculinity and a more obtuse angle with femininity. Another important landmark is the intercanthal line, an imaginary horizontal line that extends from the medial canthus of the eye to the ascending process of the maxilla. It corresponds to the area where the medial canthal tendon attaches to the ascending process of the maxilla, just anterior to the crest of the lacrimal bone.

The nasal bones are fused in the midline, and it is this area that forms the bony dorsum of the nose. The thick cephalic portion of the nasal bone articulates with the equally thick nasal process of the frontal bone situated underneath the glabella. This thick segment of bone, often termed the nasal beak, is well conceptualized by the endoscopic surgeon. It is a tough bone to fracture or drill. Just under the midline of the fused nasal bones, the superior edge of the perpendicular plate of the ethmoid articulates, forming a deep midline bony structure within the bony vault. The lateral aspects of the nasal bones articulate with the ascending processes of the maxilla and form the side walls of the nasal bridge. A portion of the maxillary ascending process itself also forms the most lateral aspect of the dorsal side wall. A fracture of the nose usually involves disarticulation or comminution of paired nasal bones, in that they are the farthest projected bones from the midface. Conversely, most rhinoplasty procedures usually involve controlled fracture and realignment of the nasal bones, given their important influence on bridge width, height, and shape of the nose. The base of the pyriform aperture is formed from the superior border of the premaxilla. The nasal spine is a small bony prominence in the midline situated at the base of the pyriform aperture. This is formed by the fusion of two halves of the premaxilla during embryogenesis (see Chapter 37). It functions as an important support strut for the cartilaginous nasal septum.

The cartilaginous components of the external nasal framework are divided into three groups: the upper lateral cartilages, the lower lateral cartilages, and the cartilaginous septum. All cartilaginous structures are paired except for the nasal septal cartilage. The paired upper lateral cartilages comprise the majority of the middle third of the nose and together with the septal cartilage, form the cartilaginous dorsum (**Figs. 38–1** and **38–3**). Although they are paired, the upper lateral cartilages are separated along the length of their midline by the septal cartilage; this configuration creates a pair of arched chambers termed the cartilaginous vaults. At their cephalic end, the upper lateral cartilages do not articulate with the nasal bones in an end-to-end manner, but are instead gently overlapped by nasal bones (**Figs. 38–1** and **38–3**). Dense fibrous tissue connects the overlapped cephalic edges of the upper lateral cartilages to the nasal bone. At this cephalic end, the upper lateral cartilages join the septal cartilage to form one single piece of cartilage, which, in turn, is overlapped by the portion of the nasal bones and extends a few millimeters posteriorly within the bony vault to articulate with the perpendicular plate of the ethmoid. The confluence of these four structures—the upper lateral cartilages, the septal cartilage, the nasal bones, and the perpendicular plate of the ethmoid—is called the keystone area and provides critical support for the nasal dorsum in the middle third of the nose. At their caudal ends, the upper lateral cartilages gently curve away from the midline, only to be overlapped again, this time by the cephalic edges of the lateral crus of the lower lateral cartilages. This junction is known as the scroll region and has both rhinometric and aesthetic significance (**Fig. 38–1**). With regard to rhinometry, early anatomical studies showed that this junction exhibits an interlocking configuration and that the degree of overlap of these cartilages influences the function of the internal nasal valve, the major determinant of air flow in the Caucasian nose. Aesthetically, the degree of this curvature determines in part the flare and fullness, or bulbousness, of the lobule. At the caudal edges of the upper lateral cartilages where they gently diverge away from the midline to meet the diverging lateral crura of the lower lateral cartilages, a small, midline, triangular space is created termed the weak triangle (**Fig. 38–3**). This area is the inferior boundary of the middle third of the nose, the caudal boundary of the dorsal subunit, and marks the area of the supratip breakpoint.

The lower lateral cartilages, also referred to as alar cartilages, traditionally have been divided into a medial and lateral crus, connected by an anatomical dome segment (**Figs. 38–1** and **38–3**). The concept of a distinct and independent middle crus has been challenged. The medial crus can be subdivided into a footplate segment and a columellar segment. Both of these segments give shape and strength to the columella. Dense fibrous tissue connects the left and right medial crura. The medial crus, columella, nasal floor, and soft tissues of the nostril rim help form the external nasal valve, a minor regulator of airflow. The anatomical dome, middle or intermediate crus, is subdivided into a lobular and domal segment. The junction of the columellar segment of the medial crus and the lobular segment of the intermediate crus determines the convexity of the columella, termed the columellar breakpoint (**Fig. 38–1**). Acute angulations in the columellar breakpoint can produce an unattractive protrusion. The domal segment is usually quite short and also frequently the thinnest, most delicate, and narrowest portion of the entire alar cartilage. Its importance in aesthetics, however, cannot be overemphasized because it is one of the major determinants of the tip configuration. The convexity of the domal segment produces a protrusion of the overlying soft tissue elements, as seen from the front view. The most anterior projecting point on this protrusion is termed the tip-defining point (**Fig. 38–2**). As stated earlier, the tip of the nose is defined as the midpoint of a line extending from one tip-defining point to the other. Thus the position of the tip is directly related to the position of the tip-defining points, which, in turn, is related to the shape of the domal segments beneath. The surface expression of the domal segment depends on three factors: its specific angulations, its position relative to the contralateral domal segment, and the thickness of the overlying soft tissue.

The lateral crus is the largest component of the lower lateral cartilage and plays a crucial role in defining the anterosuperior portion of the ala (**Fig. 38–3**). Medially, the lateral crus is contiguous with the domal segment of the middle crus, and interdomal ligaments connect this area with its contralateral counterpart. Laterally, the twin crura diverge and approach the pyriform aperture (**Figs. 38–1** and **38–3**). The aperture falls short of the pyriform edge of the maxilla, instead articulating with the first of a chain of three or four accessory cartilages called lesser alar that abut the pyriform edge of the maxilla. The caudal edge parallels the anterior half of the alar rim and contributes to its support. The cephalic edge forms a junction with the caudal edge of the upper lateral cartilage in an area known as the scroll region. It is in this region where the edges of the two cartilages interlock with one another in a curling fashion. In the nasal cavity this region contributes to the internal nasal valve. As already stated, this region has both aesthetic and functional significance.

The nasal septum is a midline structure that divides the nasal cavities into left and right halves. It is composed of both bony and cartilaginous components (**Fig. 38–4**). The most external component, however, is cartilaginous and is referred to as the quadrilateral plate because of its irregular quadrilateral shape. It projects anteriorly from the pyriform aperture and, along with the upper lateral cartilages, forms an important support structure for the dorsum of the nose from the rhinion to the supratip. Disease states adversely affect the blood supply to the septal cartilage and weaken the integrity of this structure. Collapse of the quadrilateral plate can result in a saddle nose deformity. The dorsal border of the septal cartilage at its caudal end makes an acute angle to turn ventrally. This angle is called the anterior septal angle and separates the dorsal septal border from the caudal septal border (**Fig. 38–4**). Along with the medial crura, the caudal septum provides shape and support for the columella. The anterior septal angle, in turn, is a critical support for the nasal tip. Suspensory ligaments of the tip extend from the septal angle to the cephalic edges of the lower lateral cartilages. The

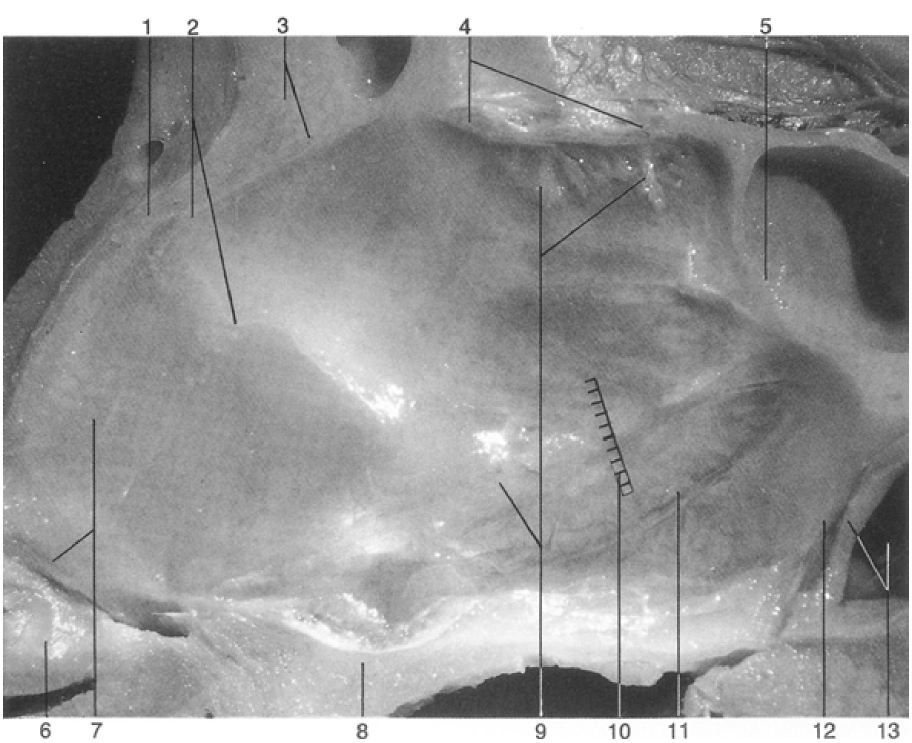


__Figure 38–4__ Bony and cartilaginous structures of the nasal septum.

cartilaginous septum forms a special relationship with the upper lateral cartilages at the level of the scroll region. From a cross-sectional standpoint, the dorsal septum makes an angle with the upper lateral cartilages and is termed the nasal valve area. This angle, in combination with the scroll region, constitutes the internal nasal valve. The optimal angle of the nasal valve should be between 10 and 15 degrees. Smaller angles tend to cause airway obstruction, especially in long Caucasian noses.

At the most ventral aspect of the caudal septal border, the septum articulates with the premaxilla on the nasal floor. The anterior nasal spine, a small bony prominence jutting off anteriorly from the premaxilla, supports the septum. The septum rests on the midline ridge of the premaxilla called the maxillary crest (**Fig. 38–4**). The crest runs anterior to posterior along the length of the floor of the nasal cavity. In the nose of people of African decent, the anterior nasal spine may be less prominent or absent. The relationship of the septum to the maxillary crest is a major factor in the straightness of the septum. Overriding of the septal cartilage over the crest or actual buckling of the cartilage near the floor of the nose significantly impedes airflow.

SOFT TISSUE AND NEUROVASCULAR STRUCTURES OF THE EXTERNAL NOSE

In addition to genetic factors, the color and texture of the skin of the nose are influenced by aging and the natural elements. As one of the most exposed areas of the body, the nose is subjected to a significant amount of ultraviolet (UV) radiation, which, through alterations of the genetic controls, promotes the growth of freckles, keratoses, and other dermatological aberrations. Whereas people of African decent have significantly more UV-protective melanin within their skin and are less inflicted with these conditions, Caucasians are at a significantly higher risk for benign and malignant neoplasm of their nasal skin. The passage of time causes an increased frequency of exposure to the elements as well as a decrease in the ability of the body to correct incurred genetic defects. As a result, atrophy of adipose and connective tissue occurs and causes the loss of definition and skin turgor. The skin of the nose tends to be thinner and more mobile in the upper half and thicker and more adherent distally. There are usually more sebaceous glands in the lower half of the nose, causing an oiliness and thickness in the skin that may limit definition. This is often the case in the non-Caucasian nose, which may have a larger and denser subcutaneous fibrofatty layer. The skin over the external nose is continuous with the skin lining of the most external portion of the internal nose, termed the vestibule. This area, from the nostril rim to the internal valve, is lined with skin that is covered with stiff hairs called vibrissae. Beneath the dermis of the skin and adherent to it is a variable layer of fibrofatty tissue. The distribution of this layer is also asymmetric, with it being thinner over the dorsum and thicker over the lobule and ala.

Beneath this layer is the fibromuscular layer, which includes the nasal subcutaneous muscular aponeurotic system (SMAS). The nasal SMAS is continuous with the SMAS of the entire face, including the platysma, and includes the facial muscles that control movement and integrity of the nose. The muscles are divided into four groups and include the elevators (muscles that shorten the nose and dilate the nostrils), the depressors (muscles that lengthen the nose and dilate the nostrils), the minor dilator, and the compressors (muscles that lengthen the nose and narrow the nostrils) (**Fig. 38–5**). The elevators are the procerus, the levator labii superior alaeques nasi, and the anomalous nasi. The depressors include the alar portion of the nasalis (dilator naris posterior) and the depressor septi. The minor dilator muscle is the dilator naris anterior. The compressor muscles are the transverse portion of the nasalis and the compressor narium minor.

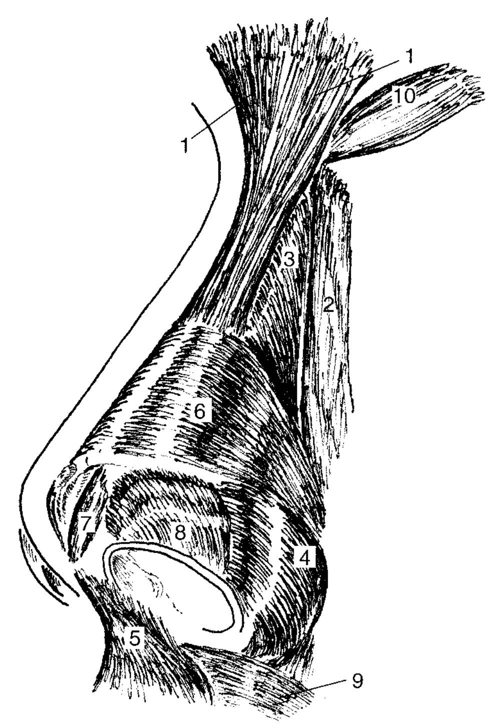


Figure 38–5 *Line drawing of the musculature of the nose. A. Elevator muscles: 1, Procerus; 2, Levator labii alaequae nasi; B. Depressor muscles: 4, Alar nasalis; 5, Depressor septi nasi; C. Compressor muscles: 6, Transverse nasalis; 7, Compressor narium minor; D. Dilator muscles: 8, Dilator naris anterior; E. Other: 9, Orbicularis oris; 10, Corrugators*

The zygomatic and temporal division of the facial nerve innervates all the muscles of the nose. The deep fatty layer runs underneath the muscles of the SMAS, and in this layer are found the major superficial blood vessels. The arteries that lie in this plane send perforating vessels superficially to supply the dermal and subdermal plexus of the overlying layers. Grossly, the branches of the external carotid artery through the facial and internal maxillary branch supply the middle and lower third of the nose, and the branches of the internal carotid artery through the ophthalmic branch supply the upper or cephalic third. It is important to remember, however, that anastomoses exist among all three arterial sources, resulting in substantial overlap. The lateral surface of the lower third of the nose including the ala and lobule is supplied by the lateral nasal branch of the angular artery; this, in turn, is the distal continuation of the facial artery (**Fig. 38–6**). This vessel also receives blood supply via anastomoses with vessels from the infraorbital and dorsal nasal arteries. Branches of the superior labial artery, another branch of the facial artery, supply the nostril sill and subnasale. The columellar artery is a distal branch of the superior labial that ascends the columella. Near the base of the columella, a septal branch of the superior labial artery is given off and supplies nearly the entire mucoperichondrium of the nasal septum.

The major blood vessel that supplies the dorsum and side walls of the nose is the dorsal nasal artery, a terminal branch of the ophthalmic artery (**Fig. 38–6**). The dorsal nasal artery perforates the orbital septum above the medial canthal tendon and descends along the side wall and dorsum of the nose to eventually anastomose directly with branches of the lateral nasal branch of the angular artery inferiorly (**Fig. 38–6**). Because the lateral nasal artery is a distal branch of the facial artery, this anatomical configuration establishes a continuous vascular conduit between the external carotid system (through the facial artery below) and the internal carotid artery system (through the ophthalmic artery above). Minor anastomoses are also made with the infraorbital artery laterally, which mainly contributes to the blood supply of the side walls. In combination with terminal branches of the columellar artery, the lateral nasal artery is the major contributor of blood to the nasal tip region. Small branches of the supratrochlear artery, a branch of the ophthalmic artery as it exits the skull at the superomedial aspect of the orbital rim, also supply the root of the nose.

The venous drainage of the nose has the same-named veins, which run parallel the arteries. These veins drain the nose via the facial vein and pterygoid plexus and through the ophthalmic veins into the cavernous sinus.

Sensation of the nose is mediated via innervations from branches of the ophthalmic and maxillary divisions of the trigeminal nerve (cranial nerve V). Grossly, branches from the ophthalmic division innervate the superior two thirds, while branches from the maxillary division mainly innervate the lower third. As with the blood vessels, substantial regional overlap exists. At the level of the radix, rhinion, and cephalic portion of the nasal side walls, the supratrochlear and infratrochlear nerves supply sensation. They are both branches of the ophthalmic division. The lower dorsum and nasal tip region is supplied by the external nasal branch of the anterior ethmoid nerve. Accompanying the artery of the same name, it emerges between the junction of the nasal bone and upper lateral cartilage. This configuration makes the external

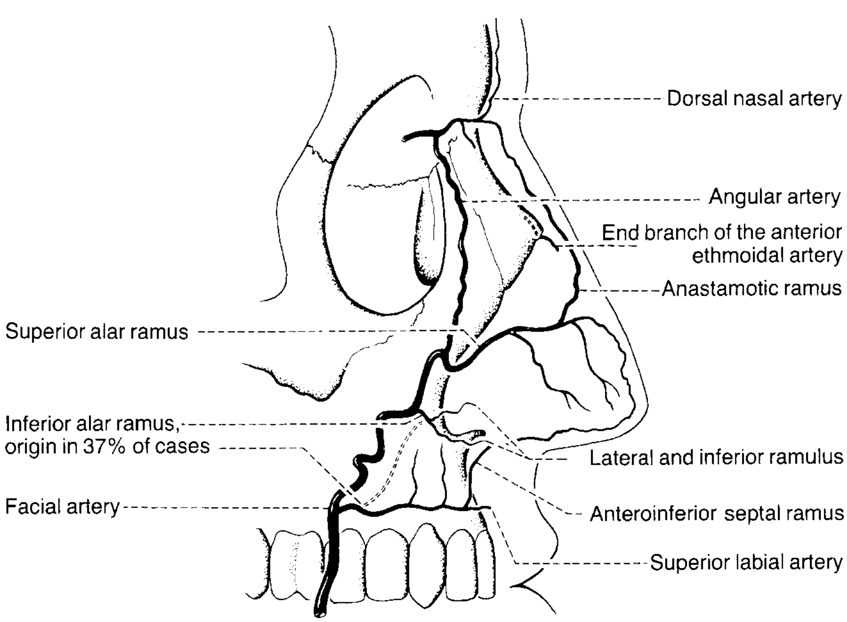


Figure 38–6 Line drawing of the key arterial vessels of the nose.

nasal branch vulnerable to injury from trauma to the dorsum of the nose, resulting in tip anesthesia. The soft tissues of the lower third of the nose, including the ala, lobule, nostril rim, subnasale, and columella, have sensory innervations through branches of the infraorbital nerve, which in turn is a branch of the maxillary division.

ENDOSCOPIC ANATOMY OF THE NOSE

The entrance to the nasal cavity is termed the vestibule, and it is lined with squamous epithelium that contains numerous thick, stiff hairs called vibrissae. The soft tissue of the nasal ala and lobule forms the covering, and the columella forms the medial wall. The medial and lateral crura provide support to the vestibule. The transition of the squamous epithelium to columnar epithelium occurs at the limen nasi, often called the white line, and marks the start of the nasal cavity.

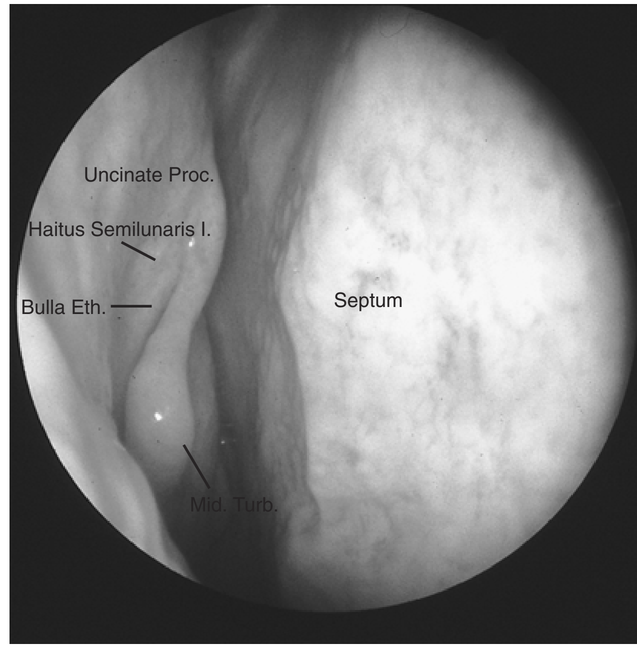


Figure 38–7 *Endoscopic view of the cadaver right nasal cavity using a 0 degree, 4 mm telescope.*

Nasal Septum

The nasal cavity is divided in the midline into two separate cavities by the septum, which is made up of a cartilaginous and a bony segment (**Figs. 38–4** and **38–7**). The cartilaginous septum is a flat plate of cartilage of irregular quadrangular shape that is fused in place by its articulation with the premaxilla inferiorly, the vomer posteroinferiorly, and the perpendicular plate of the ethmoid bone posterosuperiorly. As already stated, it also forms an important support structure for the dorsum of the nose from the rhinion to the supratip. A mixture of crossed perichondrial and periosteal fibers firmly adhere the septum to the premaxilla.

The bony septum is basically composed of two midline bones, the vomer and the perpendicular plate of the ethmoid (**Figs. 38–4** and **38–8**). The perpendicular plate forms the cephalic portion of the bony septum and is continuous with the frontal bone and the cribriform plate. Trauma to this portion of the septum carries a potential risk of injuring the cribriform plate and subsequent

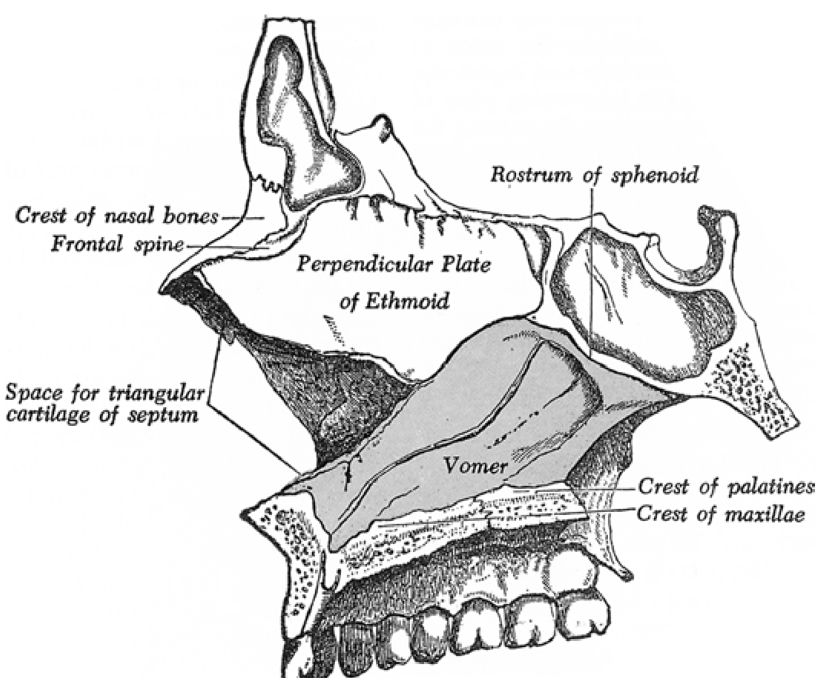


Figure 38–8 *Line drawing of the sagittal view of the nasal septum with the cartilaginous quadrilateral plate removed.*

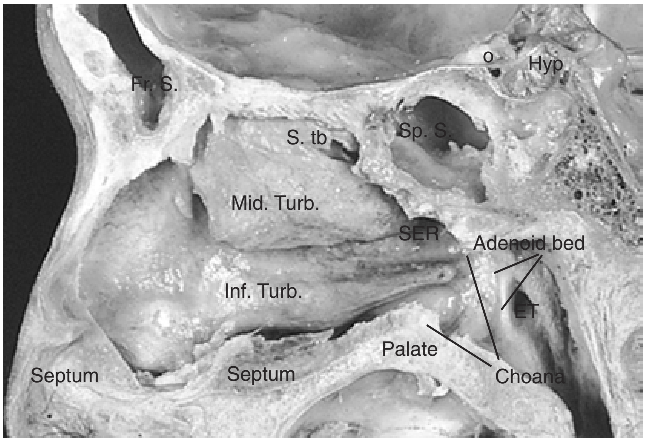


Figure 38–9 *Midsagittal section of the cadaver nose showing lateral nasal wall. ET, eustachian tube opening; FrS, frontal sinus; Hyp, hypophysis; O, optic nerve; SER, sphenoethmoid recess; Stb, superior turbinate.*

cerebrospinal fluid (CSF) leakage. Anteriorly, the perpendicular plate articulates with the undersurface of the nasal bones in the midline, caudally with the septal cartilage, and inferiorly with the vomer. The extent of contact between the perpendicular plate of the ethmoid and vomer is a function of the amount of septal cartilage that is wedged between them.

The vomer extends from the lower half of the sphenoid rostrum posteriorly to the nasal crest of the palatine bones and maxilla anteriorly (**Fig. 38–8**). It forms the lower portion of the bony septum like the keel of a ship. The most posterior aspect of the vomer forms the partitioning wall of the choana, the openings of the nasal cavity into the nasopharynx.

Turbinates

Whereas the medial wall of the nasal cavity is relatively simplistic in its shape and structure, the lateral walls of the nasal cavity are relatively complex. Three scroll-like bones termed turbinates, also called the inferior, middle, and superior conchae, project from the lateral wall in an ascending manner (**Figs. 38–9** and **38–10**). By definition, the inferior turbinate is the most inferior and is also the largest and longest of the three (**Fig. 38–10**). It is a complete and separate bone unto itself and articulates with the medial surfaces of the maxilla and the palatine bone. The long maxillary process of the inferior concha bone on its lateral aspect covers part of the maxillary hiatus in the maxillary bone and thus becomes part of the bony medial wall of the maxillary sinus. The short lacrimal process of the inferior concha articulates with the inferior aspect of the lacrimal bone and thus forms the lower part of the bony nasolacrimal canal. The posterior end of the inferior turbinate can become hypertrophied or polypoid, often projecting into the choanal space, and occasionally contributing to airway obstruction or sleep apnea. The area lateral to the inferior concha is called the inferior meatus. Both the medial wall of the maxilla and the perpendicular plate of the palatine bone form the respective anterior and posterior portions of the lateral wall of the inferior meatus. The nasolacrimal canal empties into the anterior part of the meatus. The

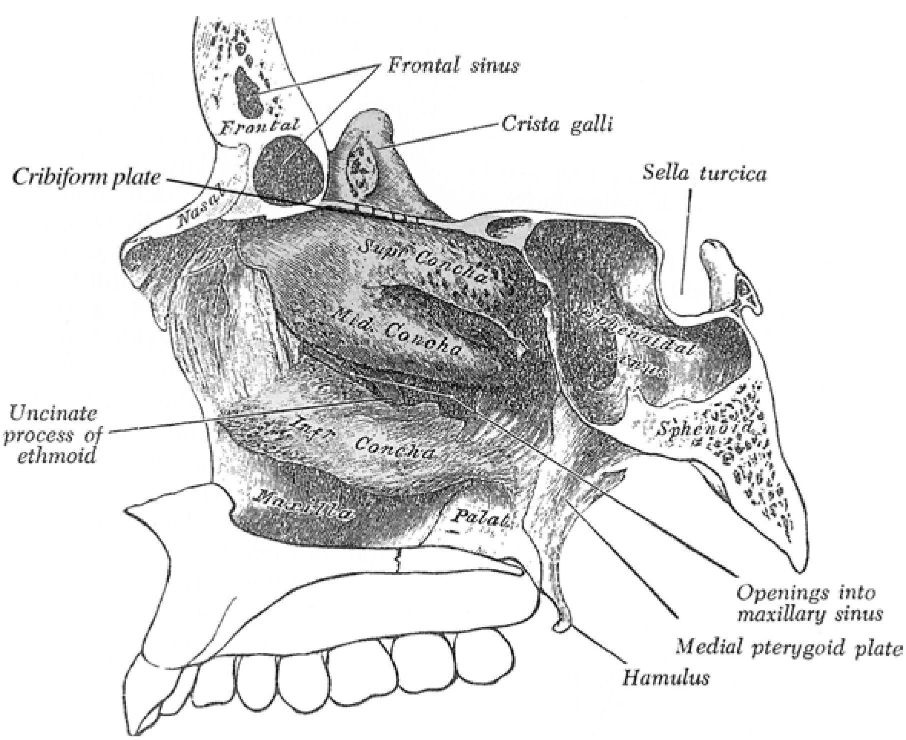


Figure 38–10 *Line drawing of the sagittal view of the lateral nasal wall with conchal bones in place.*

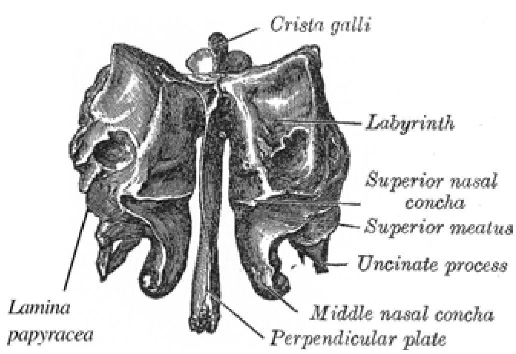


Figure 38–11 *Line drawing of the posterior view of the ethmoid bone in isolation.*

thick alveolar part of maxilla forms the lateral wall of the inferior meatus in its anterior third, and it is difficult to penetrate. The lateral wall of the inferior meatus in its middle third is formed from the thin membranous bone of the inferior concha. As stated earlier, this portion covers the maxillary hiatus, and it is easy to puncture. This fact is applied in a procedure termed an inferior antrostomy in which this thin bone is punctured to gain access into the maxillary sinus from the inferior meatus.

Unlike the inferior concha, the middle and superior conchae are part of the ethmoid bone (**Figs. 38–10, 38–11**, and **38–12**). The middle and superior conchae form the medial wall of the ethmoid bone, and they project into the nasal cavity as scrolled bones in a similar manner as the inferior concha (**Fig. 38–11**). Variably, there is a fourth concha present, called the supreme concha, which represents a third superior-most ethmoid projection in the nasal cavity. The middle turbinate, the next largest concha, is firmly attached at its anterior and posterior ends to the crests of the ascending process of the maxilla and the perpendicular plate of the palatine bone. A partitioning wall of ethmoid labyrinth called the basal lamella firmly attaches to the middle turbinate laterally in its middle third portion to the lamina papyracea. The anterior third of the middle turbinate is vertically attached to the skull base through the anterior extension of the basal lamella. As already stated, the attachment of the posterior third of the middle turbinate is to the maxillary and palatine bones. Aggressive resection of basal lamella during endoscopic sinus surgery often leads to lateralization of the anterior and midportion of the middle turbinate. The medial and lateral surfaces of the middle turbinate are convex. It is termed paradoxical when the medial surface is concave or perpendicular.

The middle turbinate overlies a space called the middle meatus that embodies two important components of the ethmoid bone: the ethmoid bulla and the uncinate process (**Fig. 38–12, 38–13**, and **38–14**). The ethmoid bulla is a thin-walled, bulging bony structure that is pneumatized by one or more ethmoid cells. It is suspended from the undersurface of the basal lamella and partially overlies the opening of the maxillary hiatus (**Figs. 38–13** and **38–14**). The uncinate process is a crescent-shaped, delicate bony plate that runs from

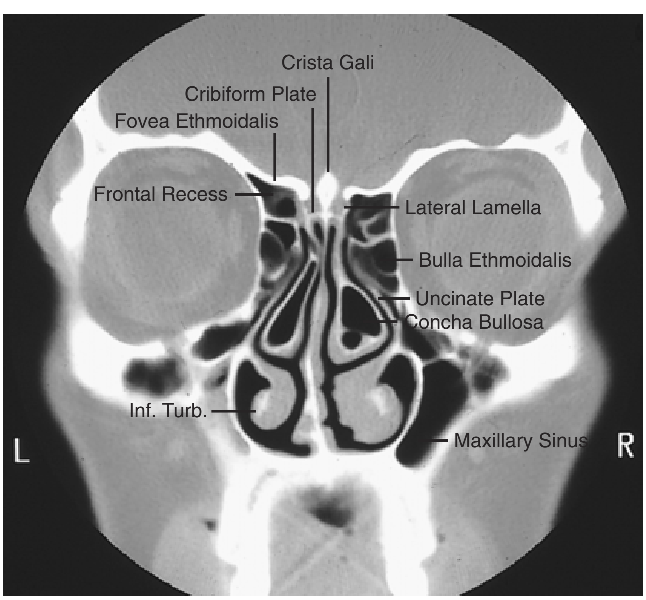


Figure 38–12 *Computerized axial tomography (CAT) radiograph of the nose. Coronal section view of the osteomeatal unit. Note bilateral concha bullosa.*

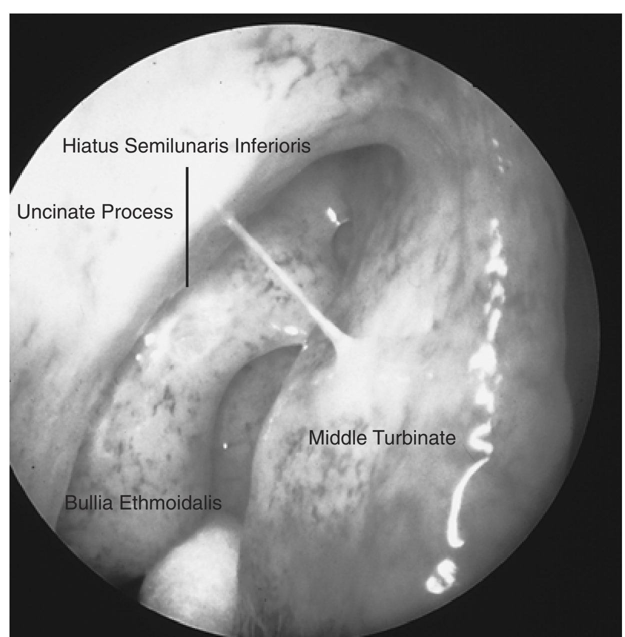


Figure 38–13 *Zero-degree endoscopic view of the specimen in* ***Fig. 38–7*** *showing right middle meatus. The middle turbinate was mobilized medially with a Q-Tip.*

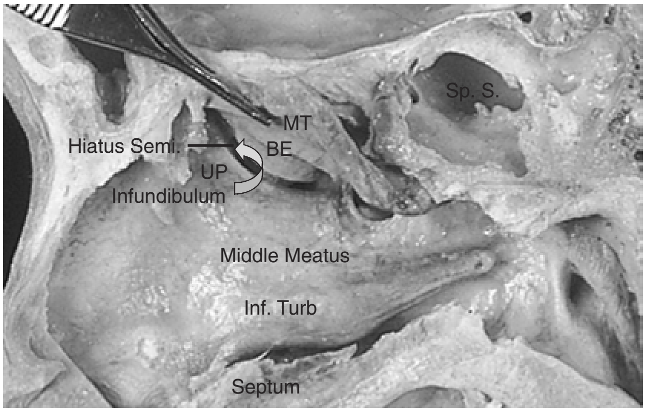


Figure 38–14 *Lateral nasal wall in* ***Fig. 38–9*** *showing the anatomical landmarks of the middle meatus. The anterior end of the middle turbinate was moved medially with forceps. The curved arrow indicates the location of the infundibulum behind the uncinate process. BE, bulla ethmoidalis; MT, middle turbinate; SpS, sphenoid sinus; UP, uncinate process.*

an area a few millimeters anterior to the bulla to its entire length, curving posteroinferiorly and slightly laterally (**Fig. 38–14**). At its posterior end it is partially fused with the inferior concha and the anterior edge of the perpendicular plate of the palatine bone. This area is defined as the posterior fontanel and is often perforated by an accessory ostium of the maxillary sinus. The blind triangular space formed by the uncinate bone anteromedially, the lacrimal bone covering the maxillary hiatus and orbit anterolaterally, and the ethmoid bulla wall posteriorly is known as the ethmoid infundibulum (**Fig. 38–14**). The contents of the maxillary, anterior ethmoid and frontal sinuses all drain into this space. The hiatus semilunaris inferioris is a two-dimensional crescent-shaped space formed by the curvature of the uncinate process and is the space into which the contents of the infundibulum drain (**Fig. 38–14**). The middle concha overlaps the ethmoid bulla and the medial wall of the uncinate. They are thus two endoscopically important landmarks of the middle meatus (**Figs. 38–13** and **38–14**).

The superior turbinate is the smallest concha. It lies superior to the posterior portion of the middle concha (**Fig. 38–11**). It covers a space called the superior meatus, and it contains the opening of the posterior group of ethmoid sinus cells.

Although the conchae are covered by mucosa in continuity with the rest of the nasal cavity, the mucosa in the area of the concha contains a plexus of large veins, which can become quite markedly engorged with blood. This feature, in conjunction with their scroll-like shape, makes them an excellent structure for promoting heat and humidity exchange as well as turbulent airflow. Thus the term *turbinate* is used to describe an individual concha and its mucosal covering as a single unit.

BONY LATERAL NASAL WALL

Aside from the conchal bones, the lateral wall of the nasal cavity is composed of a series of bones that also contribute to the structure of the midface, palate, and orbit. Several of these bones participate in the system of pneumatized air cells that form the mucosa-lined sinuses.

The maxilla bone is the most peripheral and anterior bone of the lateral nasal wall (**Fig. 38–10**). The medial surface of the maxilla has a large defect, called the maxillary hiatus, that is the opening into the maxillary sinus. The maxillary hiatus should not be confused with the natural maxillary ostium, which is a well-constructed channel in the overlying mucosa that allows drainage from the maxillary sinus. As stated, the overlying membranous bony wall of the inferior concha, the ethmoid bulla, the uncinate process, the lacrimal bone, and the palatine bone bridges the opening of this hiatus in the maxilla. The superomedial part of the maxilla is described as the ascending or frontal process. It articulates with the lacrimal bone posteriorly. The canal created by this union is the superior portion of the nasolacrimal canal and extends from the medial aspect of the orbital rim to the inferior meatus below. Thus the maxilla, the lacrimal bone, and the inferior concha all contribute to the nasolacrimal canal. The canal can be injured while removing the uncinate plate close to its articulation with the ascending process of the maxilla.

The palatine bone is composed of a perpendicular and a horizontal plate (**Figs. 38–10** and **38–15**). This is not to be confused with the perpendicular plate of the ethmoid, which is part of the bony septum. The perpendicular plate of the palatine bone covers the posterior border of the maxillary hiatus. It has articular surfaces for the articulation of the inferior and middle turbinates.

The ethmoid bone is the most complex of the lateral nasal wall bones. It consists of a horizontal plate, a vertical plate, and a bony labyrinth of air cells. The lateral wall of the bony labyrinth (sometimes called the orbital plate) is composed of a smooth lamina of thin bone termed the lamina papyracea (**Fig. 38–11**). It is, literally, a paper-thin plate. It forms part of the medial wall of the orbit. Its presence is not an anatomical constant, and when it is absent or dehiscent, infection in the ethmoid sinuses can easily spread into the orbit. Unlike the lateral wall, the medial wall of the ethmoid labyrinth is not a smooth plate. It has two or three conchal bones projecting from it that form the bony components of the middle, superior, and, when present, supreme turbinates

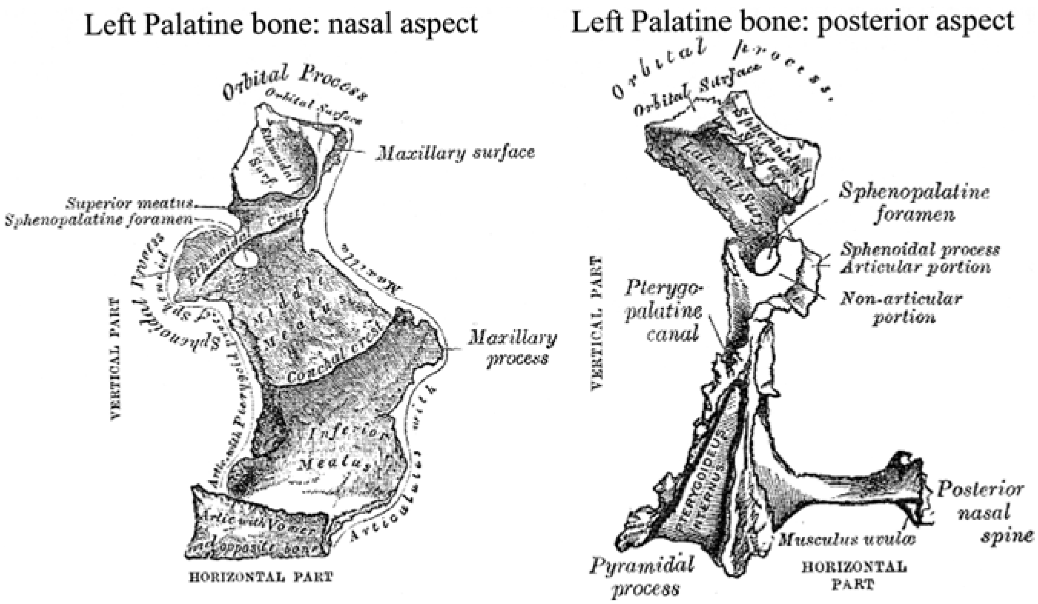


Figure 38–15 *Line drawings of the lateral and posterior views of the palatine bone in isolation.*

(**Fig. 38–11**). The narrow, perforated, thin horizontal plate of the ethmoid bone is called the cribriform plate and forms the roof of the nasal cavity (**Figs. 38–10** and **38–12**). It is connected laterally with the superior surface of the ethmoid labyrinth, also known as the fovea ethmoidalis. The fovea ethmoidalis forms the roof of the ethmoid sinus. Compared with the cribriform plate, the fovea ethmoidalis is a strong plate of bone formed by the fusion of the superior surface of the ethmoid labyrinth with the ethmoid process of the frontal bone. In the sagittal plane the fovea ethmoidalis plate is at the higher level at its anterior frontal end. In the coronal plane the sloping fovea ethmoidalis is at a higher level and is in continuity with the cribriform plate at its medial border (**Fig. 38–12**). The branches of the olfactory nerves perforate the cribriform plate and innervate the superior aspects of the nasal cavity mucosa. In this region, small fractures in the cribriform plate or a small tear in the overlying dura can result in a significant CSF leak. Because of the hidden location of the cribriform plate, endoscopic repair of CSF leaks in this area is sometimes difficult. The vertical plate of the ethmoid bone that lies in the midline and projects intracranially between the frontal lobes of the brain is called the crista galli (**Figs. 38–10** and **38–12**).

Most of the volume of the ethmoid is occupied by air-filled trabeculated cells collectively known as ethmoid sinuses or the ethmoid labyrinth. Very small, thin plates of bone partially separate the cells from one another, and each cell is lined with mucosa. This network of interconnected air cells is divided anatomically and physiologically into two groups based on the location of where they drain their contents. The division of the ethmoid labyrinth into anterior and posterior cells is made by the lateral attachment of the middle turbinate to the lamina papyracea known as the basal lamella. The anterior group of ethmoid cells is anterior and inferior to the basal lamella. The posterior group is likewise superior and posterior. The anterior group typically abuts the ascending frontal process of the maxilla, the lacrimal bone, and the nasolacrimal duct, and empties into the most superior aspect of the infundibulum. The anterior group collectively drains into the middle meatus through the hiatus semilunaris. The sinus lateralis is the most posterior of the anterior ethmoid cells and lies behind the bulla ethmoidalis, occasionally pneumatizing the attachment of the middle turbinate to the lateral wall. Despite its location, it too empties into the hiatus semilunaris inferioris. The posterior ethmoid cells collectively drain into the sphenoethmoid recess through the superior meatus (**Fig. 38–16**). A small number of posterior ethmoid sinuses may drain in the supreme meatus, if the latter is present. Because of their posterior location, they usually abut the sphenoid sinus. At variable sites, ethmoid cells extend outside the confines of the ethmoid bone (extramural extensions) into the surrounding bony structures. For example, extension and pneumatization of the anterior ethmoid cells into the frontal process of the maxilla or any bony structure anterior to the frontal sinus opening results in agar nasi cells. Lateral extension of the anterior or posterior ethmoid cells that results in pneumatization of the orbital floor is called Haller's cells. Posterior extension of ethmoid cells into the lateral aspects of the sphenoid sinus is known as Onodi cells. Pneumatization can also take place medially in the

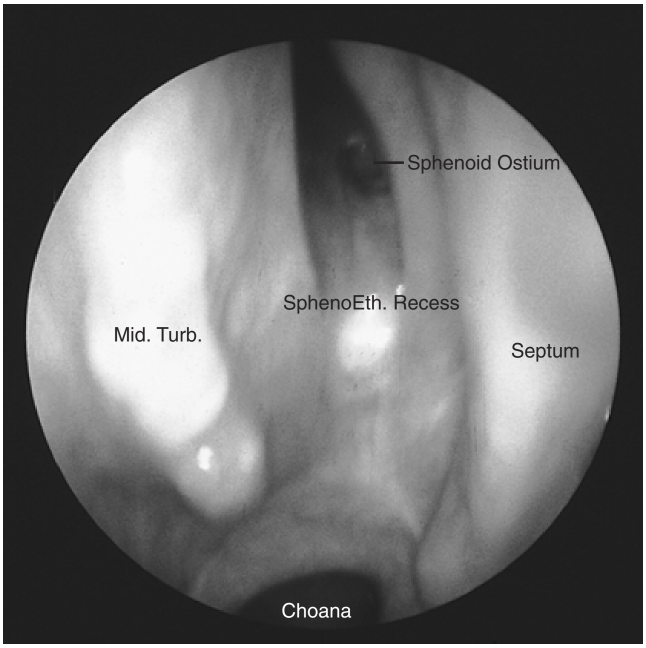


Figure 38–16 *Thirty-degree endoscopic view of the specimen in* ***Fig. 38–7*** *showing the right sphenoethmoid recess. The endoscope was advanced into the posterior end of the nasal cavity.*

bone of the middle turbinate known as a concha bullosa (**Fig. 38–12**). Likewise, extensions of ethmoid cells into the uncinate, palatine, and septum are also known.

The ethmoid contains a variable number of sinuses. Usually there are two to eight anterior sinus cells and one to five posterior sinus cells. On average, the total amount of about 10 to 15 cells per side exists with a total volume of ~15 mL. The cells of the anterior group tend to be smaller than those of the posterior group.

ROOF OF THE NASAL CAVITY

The roof of the nasal cavity is narrow in anteriorly and wide posteriorly portions. The upper lateral cartilage, the nasal bone, the frontal bone, the cribriform plate of the ethmoid bone, and the anterior wall of the sphenoid form the roof of the nasal cavity in the midsagittal section. Lateral continuity of the cribriform plate into the fovea ethmoidalis forms the lateral extension of the roof. It also forms part of the anterior skull base (**Fig. 38–12**).

The intranasal portion of the frontal bone has a central nasal spine, which articulates with the perpendicular plate of the ethmoid and plays an important role in the keystone area of the nasal dorsum. The location of the frontal sinus is within the pneumatized portion of the frontal bone (**Figs. 38–8, 38–9**, and **38–10**). The access of the frontal sinus to the nasal cavity is at the inferior aspect of the frontal bone. It is near the most anterior points of articulation with the anterior ethmoid air cells. The frontal sinus ostium is in the posteromedial part of the frontal sinus floor, and it is usually at the lowest point of that floor. Both ostia open into the pre-ethmoid chamber called the frontal recess that eventually drains into the infundibulum. The frontal sinus is divided by a complete bony septum located in the midsagittal plane.

The orbit and the orbital contents, the anterior ethmoid cells, and the nasal cavity border the frontal sinus inferiorly. Occasionally, an anterior ethmoid air cell will extend into the frontal bone to form an accessory sinus called the bulla frontalis.

The horizontal cribriform plate is the medial part of the roof of the nasal cavity (**Figs. 38–10** and **38–12**). It is divided in the midline by the cristi galli superiorly and the perpendicular plate inferiorly (**Fig. 38–11**). Along with the lateral part of the fovea ethmoidalis, the roof of the nose is a bony partition separating the ethmoid labyrinth from the olfactory bulb and frontal lobes of the brain. As part of the ethmoid complex, the cribriform plate is in direct continuity with the perpendicular plate of the ethmoid. Unlike the fovea ethmoidalis, the cribriform plate is specialized in that the olfactory bulb, which sits atop in the anterior cranial fossa, receives its sensory nerve fibers through tiny perforations in the substance of the plate. The location of the olfactory nerve to the ethmoid air cells accounts for the symptoms of hyposmia or anosmia that accompany severe ethmoid disease such as polyposis, esthesioneuroblastoma, or inverted papillomas.

FLOOR OF THE NASAL CAVITY

Like the structure of the nasal septum, the bony framework of the floor of the nasal cavity is relatively simplistic. It is composed of horizontal shelves that also serve as the hard palate of the oral cavity. The premaxilla, that portion of the maxilla that is anterior to incisive foramen, forms the most anterior segment of the floor of the vestibule (**Figs. 38–8, 38–9**, and **38–10**). The incisive foramen is a small defect in the anterior palate that results from the fusion of the primary palate (the premaxilla) with the lateral shelves of the secondary palate (the maxilla). The palatine process of the maxilla forms the next segment and composes approximately three fourths of the bony floor of the nasal cavity. This flat, horizontal process of the maxilla meets its contralateral counterpart in the midline with the maxillary crest forming the bony junction of the two. It is important to note, however, that although the premaxilla and maxilla form from two distinct cartilaginous structures embryonically, they eventually merge and form one bone at maturity. The horizontal plate of the palatine bone forms the final segment of the floor of the nasal cavity (**Fig. 38–15**). It too meets in the midline and contributes to the maxillary crest, which supports the nasal septum above it.

POSTERIOR NASAL WALL

The sphenoid sinus occupies the superior portion of the posterior nasal cavity centrally. Located within the sphenoid bone, the sphenoid sinus is an enlarged cavity separated by a complete or incomplete bony septum in the midsagittal plane (**Figs. 38–8, 38–9, 38–10**, and **38–17**). The two cavities that are formed often exhibit some asymmetry. Occasionally, smaller incomplete bony septations may exist within the two main cavities. The sphenoid sinuses drain directly into the sphenoethmoid recess through a small opening or sphenoid ostium on the anterior wall of the sphenoid. The sphenoethmoid recess is a narrow space between the anterior wall of the sphenoid sinus and the posterior ends of conchae (**Fig. 38–16**). The ostium may be as high as 14 mm above the roof of the choana. Occasionally, this does not allow for efficient clearing of fluid from the sinuses. The location of the ostium can be approximated by measuring a distance of 7 cm from the nostril rim at a thirty-degree angle from the anterior nasal spine. It is approximately 1 cm from the roof of the choana.

Because the roof of sphenoid and ethmoid bone is located in a central position in the skull, several important intracranial anatomical relations exist. They are the frontal lobes of the brain, the hypophysis, and the olfactory tract. The optic chiasm is located anterosuperior to the sphenoid sinus. Dehiscence in the superior wall of the sphenoid sinus can thus allow for the spread of infection from the sinus to critical structures within the brain. Posteriorly, where the wall is very thick, lie the pons and basilar artery. Laterally, a thin wall separates the sphenoid from the ophthalmic division of the trigeminal nerve, the cavernous sinus, the internal carotid artery, the abducens nerve, and the maxillary division of the trigeminal nerve.

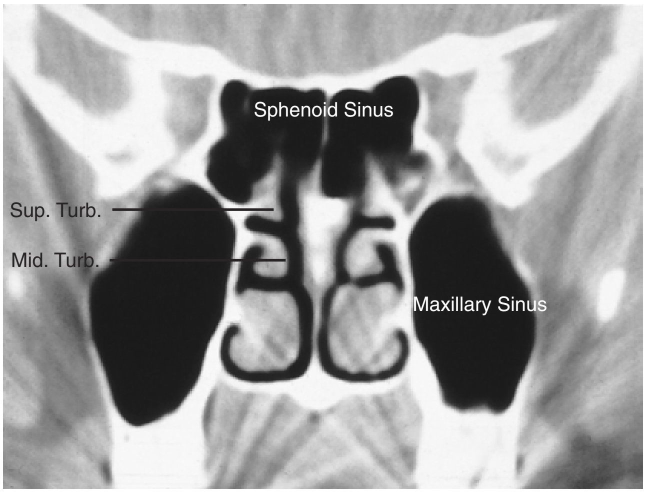


Figure 38–17 *CAT radiograph of the nose. Coronal section view of the sphenoid sinus.*

An aperture of the nasal cavity at its posterior lower end is known as the choana (**Figs. 38–9, 38–14**, and **38–16**). It is defined as the posterior opening of the nasal cavity into the superiormost aspect of the nasopharynx. It is bounded by the sphenoid sinus superiorly, the posterior end of the middle turbinate laterally, the distal edge of the horizontal plate of the palatine bone inferiorly, and the most distal edge of the vomer medially. Two important anatomical relations are the eustachian tube orifice laterally and the adenoid tissue bed posteriorly (**Fig. 38–9**). Polyps, the hypertrophied posterior end of the inferior turbinate, or nasal masses that extend out of the choana posteriorly into the nasopharynx can obstruct the eustachian tube orifice and cause middle ear disease. Likewise, adenoid tissue hyperplasia or nasopharyngeal tumors can obstruct the choana and perturb airflow through the nasal cavity.

SOFT TISSUE AND NEUROVASCULAR STRUCTURES

The entire nasal cavity, aside from the vestibule, is lined with pseudostratified columnar ciliated (respiratory) epithelium. The transition between the stratified squamous epithelium of the vestibule and the mucosa that lines the rest of the nasal cavity is called the limen nasi. However, the distribution of specialized cells and structures within the mucosa is not homogeneous. Instead, several discrete areas of specialized function exist throughout. For example, the roof of the nasal vault, the superior aspect of the septum, and the superior turbinates contain specialized sensory olfactory epithelium. Within this epithelium lie sensory neurons that send olfactory information to the brain via the olfactory nerves and olfactory bulb. At certain locations within the respiratory epithelium exist goblet cells, the main function of which is to produce a mucous layer consisting of mucoglycoproteins and acid mucopolysaccharides. In addition, the mucosa covering the conchal bones and portions of the floor is perfused with a system of capillary vessels and venous sinusoids that can engorge with blood in response to β–adrenergic stimulation, resulting in nasal congestion, and constrict with α-adrenergic stimulation, resulting in decongestion. Throughout the mucosa, however, inflammatory cells can be found that will elicit a rapid release of inflammatory mediators from mast cells, basophils, and other leukocytes in response to chemical or microbial insult.

The fifth cranial nerve (trigeminal) provides the sensory innervations of the nasal mucosa through its first (ophthalmic nerve) and second (maxillary nerve) divisions. The ophthalmic branch, entering the orbit through the superior orbital fissure as the nasociliary nerve, gives off the anterior ethmoidal nerve as a distal branch that enters the nasal cavity through a foramen at the level of the cribriform plate (**Fig. 38–18**). It innervates the anterior and midsuperior areas of the lateral and septal walls via the lateral internal and medial internal branches of the anterior ethmoidal nerve, respectively.

The major sensory innervations in the nasal cavity come from the maxillary division of the trigeminal nerve. It exits the skull base through the foramen rotundum to enter the pterygopalatine fossa, where it passes through the pterygopalatine ganglion to enter the nasal cavity via the pterygopalatine foramen as the lateral and medial nasal branches of the pterygopalatine nerve (**Fig. 38–18**). The lateral branch supplies the majority of the middle and posterior portions of the lateral nasal wall. The medial branch innervates the majority of the middle and posterior septum and has a distal branch termed the nasopalatine nerve, which sends fibers through the incisive foramen on the nasal floor to innervate a limited area on the undersurface of the hard palate. The greater and lesser palatine nerves are two other branches of the pterygopalatine nerve that pierce the soft palate via the greater palatine foramen. The greater palatine nerve innervates the majority of the mucosa on the undersurface of the hard palate and anastomoses with the nasopalatine nerve through the incisive foramen. The lesser palatine nerve innervates the mucosa of the soft palate and the uvula.

Parasympathetic and sympathetic innervations to the secretory glands of the nasal mucosa as well as to the specialized vasculature of the turbinates arrive via the pterygopalatine (also known as sphenopalatine) ganglion. The pterygopalatine ganglion is by definition a parasympathetic ganglion because the cell bodies of the postganglionic parasympathetic neurons reside in the ganglion and synapse with their preganglion counterparts. The cell bodies of the preganglionic fibers originate from the lacrimal nucleus in the brainstem and reach the pterygopalatine ganglion via the nervus intermedius division of the facial nerve and the greater superficial petrosal nerve.

Sympathetic innervations arrive from postganglionic fibers that pass through the pterygopalatine ganglion without synapsing. They reach the ganglion via the internal carotid plexus, the deep petrosal nerve, and the nerves of the pterygoid canal.

The first cranial nerve provides special sensory innervations to the olfactory epithelium. The fibers of the olfactory nerve originate as the central processes of the olfactory receptor nerve cells in the mucosa of the upper surface of the superior concha, the sphenoethmoid recess, and the corresponding areas on the nasal septum and roof. Bundles of neuron fibers merge to form central nerve fibers named olfactory nerves. Olfactory nerves along with the arachnoid and dural sheath pierce the cribriform plate of the ethmoidal bone and end in the olfactory bulb in the anterior cranial fossa. The olfactory bulb is connected to the olfactory area of the cerebral cortex by the olfactory tract.

The arterial blood supply to the nasal cavity originates from the internal maxillary and facial branches of the external carotid artery and from the ophthalmic branch of the internal carotid artery, from which the anterior and posterior ethmoidal arteries originate. The anterior

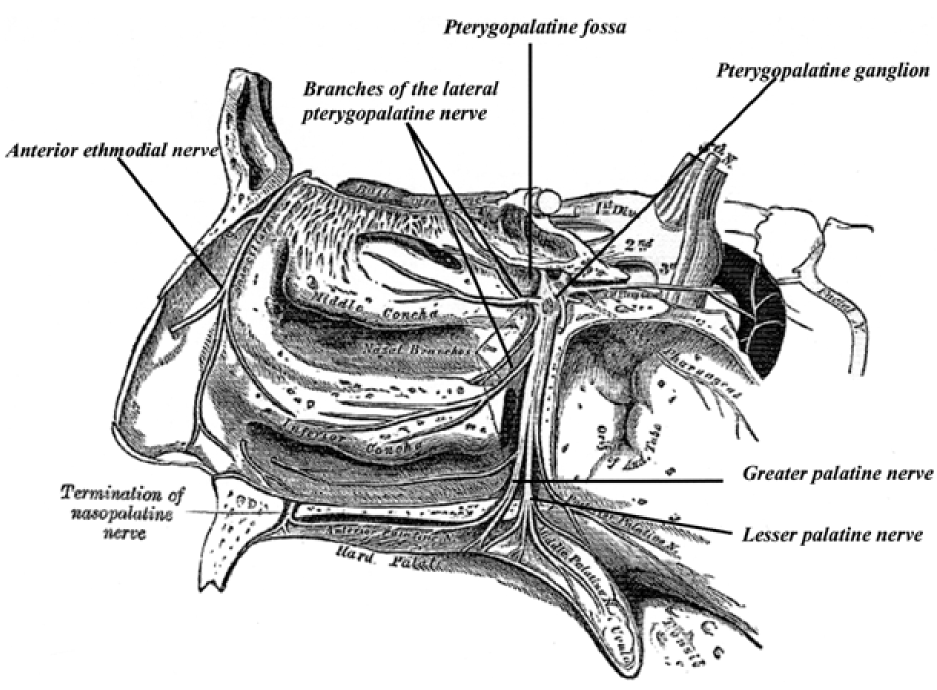
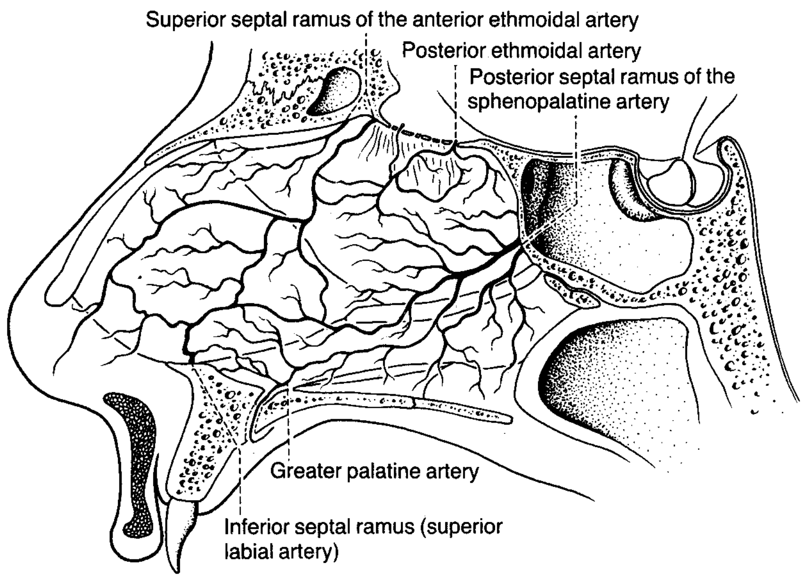


Figure 38–18 Line drawing of the sagittal view of the lateral nasal wall indicating the approximate location of the major sensory nerves.

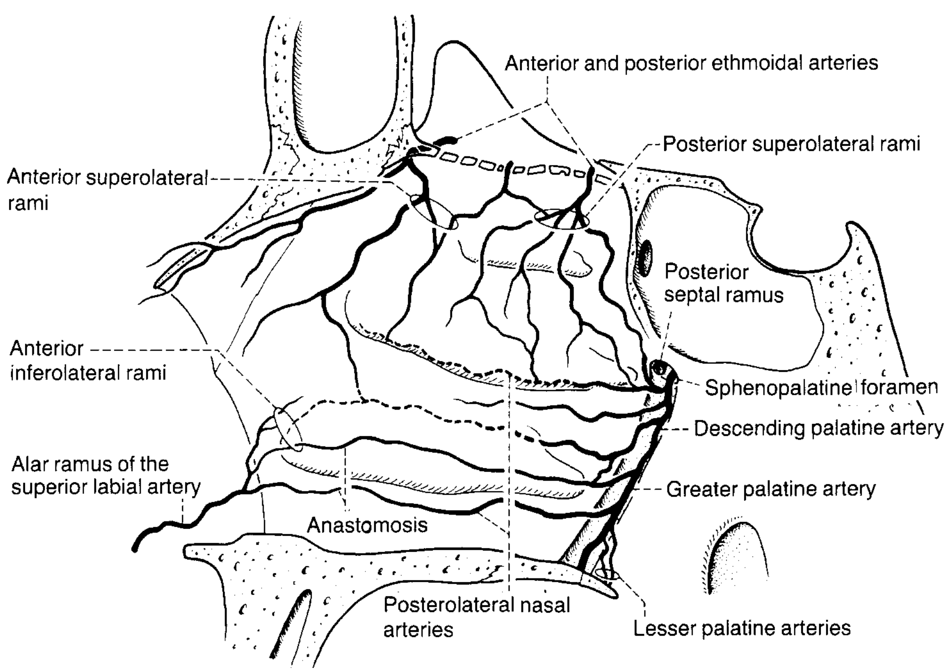
facial vein, the sphenopalatine vein, and the ethmoidal vein supply the venous drainage of the nose. The latter two veins eventually drain into the cavernous sinus.

The anterior septum is supplied mainly by the septal branch of the superior labial artery inferiorly (a branch of the facial artery), the medial internal nasal branch of the anterior ethmoidal artery superiorly (a branch of the ophthalmic artery), and the distal branches of the sphenopalatine artery posteriorly (a branch of the internal maxillary artery) (**Fig. 38–19A**). The three blood supplies merge in a region in the anterior septum known as Little's area to form a rich plexus of arteries known as Kiesselbach's plexus. Due to the richness of blood vessels, this area is often the site of anterior septal bleeds. The posterior septum is supplied mainly by the septal branch of the posterior ethmoidal artery superiorly and the proximal branches of the posterior septal branch of the sphenopalatine artery. Small branches of the sphenopalatine artery travel through the incisive foramen and supply a limited area of the undersurface of the nasal floor (the roof of the oral cavity) (**Fig. 38–19A**).

The lateral nasal wall is perfused in a similar pattern as that of the septum (**Fig. 38–19B**). The anterior area, from the midportion of the middle turbinate to the vestibule, is supplied mainly by distal branches of the angular artery inferiorly (from the facial artery) and the lateral internal nasal branch of the anterior ethmoidal artery superiorly (from the ophthalmic artery). As opposed to



A



B

Figure 38–19 *Line drawings of the sagittal views of the **(A)** medial and **(B)** lateral nasal walls indicating the approximate location of the major arteries.*

the septum, however, the posterior lateral wall is mainly supplied only by the sphenopalatine artery and not the posterior ethmoidal artery. Because it covers such a large area, posterior nasal bleeding is often the result of a perforation in this lateral nasal branch of the sphenopalatine artery. The lateral nasal branch of the posterior ethmoidal artery supplies a limited area in the vicinity of the superior turbinate and sphenoethmoid recess.

The main trunk of the sphenopalatine artery lies within the pterygopalatine (also known as sphenopalatine) fossa in the posterior aspects of the nasal cavity and enters the nasal cavity through the sphenopalatine foramen. As already stated, its branches include the posterior lateral and posterior septal branches. It also has a descending branch, call the descending palatine artery, that remains submucosal and pierces the soft palate through the greater palatine foramen to give rise to the greater palatine and lesser palatine arteries (**Fig. 38–19B**). The greater palatine artery supplies the majority of the undersurface of the hard palate and anastomoses with vessels of the posterior septal branch of the sphenopalatine artery through the incisive foramen (**Fig. 38–19A**). The lesser palatine artery supplies the soft palate and the uvula.

The lymphatics of the nasal cavity drain anteriorly through the vestibule and into the lymphatic vessels of the upper lip. Posteriorly, the lymphatics are larger and drain into the deep cervical lymph nodes. The majority, however, pass into a plexus in front of the eustachian tube, where they join the lymphatics from the upper pharynx and middle ear to pass into the retropharyngeal nodes.

Suggested Readings

Denecke HJ, Meyer R. Plastic Surgery of the Head and Neck: Corrective and Reconstructive Rhinoplasty. Vol 1. New York: Springer-Verlag; 1976

Goss CM. Gray's Anatomy. Philadelphia: Lea & Febiger; 1973

Griesman BL. Muscles and cartilages of the nose from the standpoint of typical rhinoplasty. Arch Otolaryngol Head Neck Surg 1944;39:334

Lang J. Clinical Anatomy of the Nose, Nasal Cavity and Paranasal Sinuses, trans. Stell PM. Stuttgart: Georg Thieme Verlag; 1989

McCullough EG. Nasal Plastic Surgery. Philadelphia: Saunders; 1994

McMinn RMH, Hutchings RT, Logan BM. Color Atlas of Head and Neck Anatomy. Chicago: Year Book Medical Publishers; 1981

Mehta D. Atlas of Endoscopic Sinonasal Surgery. Philadelphia: Lea & Febiger; 1993

Sheen JH, Sheen AP. Aesthetic Rhinoplasty. 2nd ed. St. Louis: Mosby; 1987

Stammberger H. Functional Endoscopic Sinus Surgery. Philadelphia: BC Decker; 1991

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. The depression on the face that marks the approximate location of the nasofrontal angle is called the
   A. Rhinion
   B. Radix
   C. Glabella
   D. Nasion
2. Facts about the weak triangle include all of the following except
   A. Its borders include margins of the medial and lateral crus.
   B. It is an unimportant structure in cosmetic surgery.
   C. It forms the inferior boundary of the middle third of the nose.
   D. Boundaries include the caudal edge of the lower lateral cartilages.
3. The following are true facts about the nasal septum except
   A. It is composed of bony and cartilaginous components.
   B. It is a key component of the internal nasal valve.
   C. It forms a support structure from the nasion to the rhinion.
   D. It supports the nasal tip.
4. Trauma to which bony structure causes the highest risk of cerebrospinal fluid leakage?
   A. Vomer
   B. Perpendicular plate of the ethmoid
   C. Perpendicular plate of the palatine bone
   D. Premaxilla
5. Penetration of which structure allows infection to spread from the paranasal sinuses to the orbit?
   A. Basal lamella
   B. Fovea ethmoidalis
   C. Uncinate process
   D. Lamina papyracea
6. Anterior ethmoidal artery ligation will best control which bleeding situation?
   A. An anterior septal bleed only
   B. A posterior septal bleed only
   C. An anterior lateral wall bleed only
   D. A posterior lateral wall bleed only
   E. An anterior septal and lateral wall bleed only
   F. A posterior septal and lateral wall bleed only
   G. All of the above

Chapter 39

NASAL AND PARANASAL SINUS PHYSIOLOGY

Erich P. Voigt and David R. Edelstein

One of the most overlooked areas in medicine is the physiology of the nose and paranasal sinuses. With the expanding popularity of new surgical approaches and endoscopic instruments, there has been an increasing need to understand how the nasal cavity functions. Although more than 200,000 sinus surgeries are performed each year in the United States alone, there is a great disparity between the paucity of biological knowledge that we possess in this field and the volume of surgical interventions being performed. Traditionally, the nose was perceived to be a simple, cosmetic conduit to the throat and lungs. Until the late 19th century, there was no clear understanding of the function of the nose. The use of rhinomanometers, photographic-compatible endoscopes, and new tools to test for allergies and ciliary dysfunction has helped to rekindle an interest in studying the physiology of the nose. It is our hope that the current enthusiasm, which otolaryngologists have demonstrated in performing surgery on the nose, will lead to great advances in our understanding of the physiology of the nasal cavity.

The nose and paranasal sinuses serve many important functions. These include exchanging air, transporting mucus, filtering noxious particles, warming and humidifying, providing a bumper from injury, providing protection from foreign organisms, olfaction, and serving as a resonator of sound. Despite inclement weather, extreme changes in the external environment, illness, or trauma, the nose must continue to function at a constant level or the respiratory tract and body will suffer. There is a great deal about the nose that we do not completely understand. Much of what we know about this organ has been inferred from our knowledge of the lower respiratory tract.

This chapter will describe, in brief, the basic physiology of the nose and paranasal sinuses. We will discuss airflow patterns, the histology of nasal epithelium, cilia, ciliary movement, mucus, mucociliary clearance of the sinuses, neuronal influence in the nose, nasal reflexes, immune function of the nose, and tests of nasal function. Nasal and paranasal sinus embryology, anatomy, and olfaction will be discussed in separate chapters of this text.

AIRFLOW PATTERNS

Approximately 10,000 liters of air per day pass into the lower respiratory tract via the nasal passages and the mouth. Each of these routes serves as a variable resistor with multiple baffles and valves. In the nose, these are the nasal alae, vibrissae, valve area, turbinates, mucosal vascular tissue, nasal passage, and nasopharynx. Changes in each of these components vary with inspiratory and expiratory movements of air. The nose is not a static area because it is always changing in response to environmental, hormonal, and age-related developments. This section describes these basic factors.

Infants, by nature, are obligate nasal breathers. An infant born with complete bilateral choanal atresia presents an immediate respiratory emergency. During early childhood, the relationship of the relatively small nasopharynx with the palate and epiglottis changes in configuration, thereby making the individual both an oral and nasal breather. During this period, the nasal cartilages and midface develop, which changes some of the relationships in the nose. By early adulthood, the nose is the preferable route for respiration. The nose is the more efficient route because of its ability to provide a type of natural positive end-expiratory pressure (PEEP) and temperature and humidity control. Individuals with nasal obstruction are significantly more likely to suffer from snoring and may have more frequent episodes of apnea and hypopnea, indicative of severe sleep-disordered breathing. The effects of aging on this area are not well known, except to note that the cartilages and vascular control of the nose vary, which changes the efficiency of the nose and its ability to pass air.

Anatomy

The medial and lateral walls of the nose are probably the most important part of the airflow of the nose. The best airflow laterally is over the inferior turbinate and through the middle meatus. Air tends to hit the anterior portion of the inferior and middle turbinates and is directed posteriorly between them. Therefore, the anterior ethmoidal area is very important for proper airflow. Ethmoidal polyps will cause significant obstruction to anterior nasal airflow. The high velocity in the front of the nose tends to help larger particles collect in the vibrissae and along the tougher squamous mucosa of the nasal vestibule. Once air passes through the nasal valve area, the cross-sectional area greatly increases, and the velocity falls rapidly. The significant decrease in velocity coupled with the viscous retardation of air by the large surface area gives rise to turbulent flow. In addition, there is a small amount of turbulent flow to the roof of the nose, which probably explains the physiology of the sniff and the route for smells to be perceived by the olfactory receptors at the roof of the nasal cavity. The medial wall airflow pattern is along the floor or adjacent to the medial turbinates. Septal deflections may significantly change these relationships.

The nasal valve can dramatically change the airflow rates. The nasal valve, as defined by Cole (1993), consists of the anterior bony cavum containing inferior turbinate erectile tissue, lateral wall erectile tissue anterior to the turbinate, and septal erectile tissue, supported respectively by compliant alar tissues and the rigid cartilaginous septum. The average valve area changes from 90 mm^2 to a thin passage of 30 mm^2 during normal respirations. This area is the narrowest passage in the upper respiratory tract, causing more than half of the total nasal resistance to respiration in normal subjects. The velocity of airflow at this point is the fastest in the entire airway. The extrinsic and intrinsic muscles can change these relationships. Every rhinoplasty surgeon should know that excessive excision of the upper and lower lateral cartilages may cause valve collapse and depress nasal respiration. The air-starved patient, such as the asthmatic or chronic bronchitic, will frequently have nasal flaring in an attempt to minimize the effect of this area on total airflow. Cole (1993) notes that the greatest linear velocities and differential pressures in the upper airways are found in the nasal valve space. Portugal demonstrated that the Breathe-Right device, which is a small adhesive band with two parallel plastic strips applied across the middle third of the nasal dorsum, increased the cross-sectional area at the nasal valve by 21%, and resulted in a 27% decrease in nasal airway resistance in Caucasian subjects.

The respiratory lining of the nose has the ability to expand dramatically, which also will change the airflow. Overall, the average resistance in the nose tends to remain the same over time. Nevertheless, exercise, pregnancy, emotional states, and medications may cause major shifts in resistance due to variations in the vascularity of the nasal lining. The most common changes are in the nasal cycle, when the lining in one nasal passage swells with blood, while the opposite side remains dormant. This is believed to allow one side to encourage more humidification and adjust to temperature changes, while the opposite side serves to maintain airflow.

The back of the nose connects with the nasopharynx, where the two passages combine into one. The normally turbulent airflow of the nose is transformed into a linear flow pattern. The presence of adenoidal tissue can change the normal posterior choanal airflow. Swelling of the posterior portion of the inferior turbinates may also modify the nasopharyngeal flow patterns.

Particles in inspired air larger than 3 μm have a maximum deposition in the anterior part of the nose, at the nasal valve area. Particles smaller than 3 μm and larger than 0.5 μm are filtered by the nasal mucosa and transported by cilia propulsion to the nasopharynx. The filtration for particles smaller than 0.5 μm is low; these particles seem to pass easily into the lower airway.

The receptors and neural pathways involved in the common symptom of nasal blockage are of particular interest. Studies to date suggest that the sensation of nasal patency may be related to the temperature of the nasal passages. Inspired air cools the nasal lining on inspiration, and this may be detected by thermoreceptors in the mucosa. Studies have demonstrated that with no change in objective measured patency, sensation of nasal airflow is decreased by the injection of local anesthetic through the buccal mucosa above the root of the upper incisor teeth, and enhanced by breathing menthol vapors through the nose or by the respiration of cooler air.

MICROSCOPIC ANATOMY

Epithelium

The nose and paranasal sinuses are lined by the respiratory epithelium, which is composed of ciliated and nonciliated pseudostratified columnar cells (which are all covered with microvilli), basal pleuripotential cells, and goblet cells (**Figs. 39–1** and **39–2**). Seromucinous glands are found in the submucosa and are important for mucus production. Cilia line most of the upper and lower respiratory tract, the sinuses, middle ear, and eustachian tube. Cilia are one of the oldest methods of propulsion in the body. These cells also serve to help with the clearance of foreign particles and to direct fluid movement. The average size of cilia is 25 μm. The average nasal cilium is 5 to 7 μm long and 0.3 μm thick. The normal nasal cell has 200 to 300 cilia with 150 microvilli.

Cilia

Cilia have a specific organization and orientation. In general, there are nine pairs of micro tubules surrounding a central pair (**Fig. 39–3**). Spokes radiate from the central fibers. There are three important types of bonds in this structure. First are the nexin links, which are elastic doublets and act as small bridges. These help to stabilize the cells. Second, are the dynein arms, which help the nine doublets to slide in concert. Third, there are spokes that keep the cilia from breaking apart.

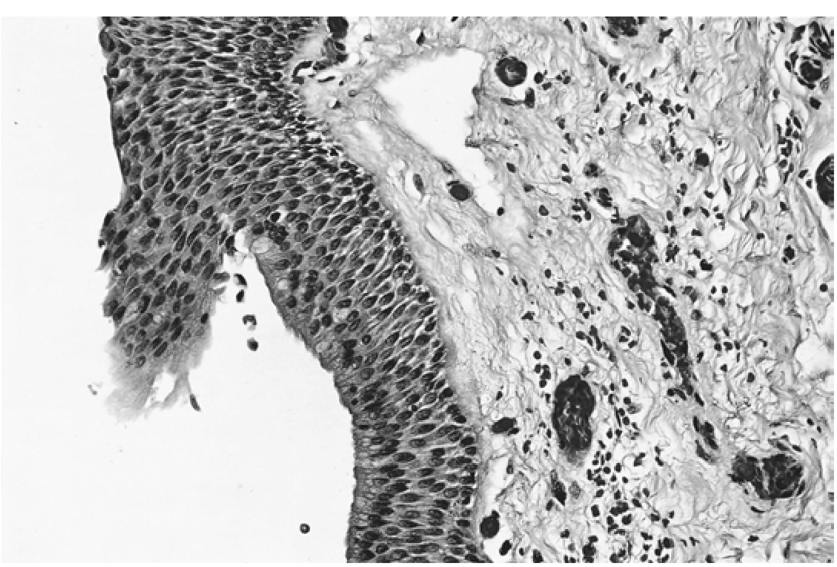


Figure 39–1 *Respiratory epithelium, normal nasal mucosa, hematoxylin-eosin stain. Note pseudostratified ciliated columnar epithelium with many goblet cells overlying highly vascular stroma with sparse infiltrate. (From Anand VJ, Pange WR. Practical Endoscopic Sinus Surgery New York: McGraw-Hill; 1993. Reproduced with permission from the McGraw-Hill Companies.)*

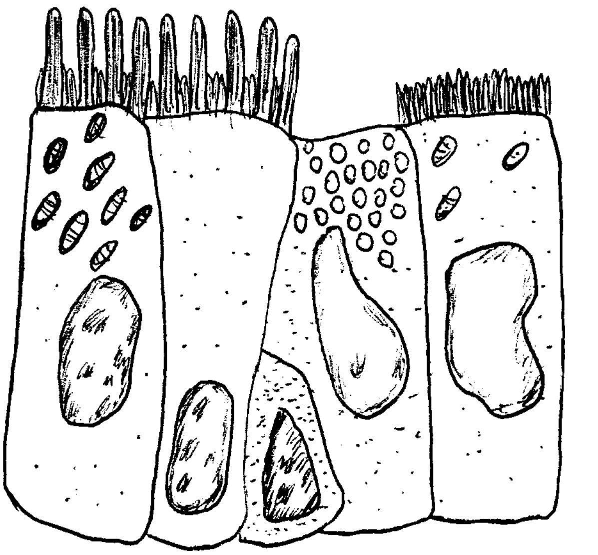


Figure 39–2 *Respiratory epithelium. The cells of the nasal epithelium: pseudostratified columnar epithelium. All cells rest on the basement membrane, but not all cells reach the surface.* ***(A,B)*** *Ciliated columnar cells.* ***(C)*** *Basal cells.* ***(D)*** *Goblet cells with mucus granules.* ***(E)*** *Nonciliated columnar cells (note microvilli). (From Anand VJ, Pange WR. Practical Endoscopic Sinus Surgery. New York: McGraw-Hill; 1993. Reproduced with permission from the McGraw-Hill Companies.)*

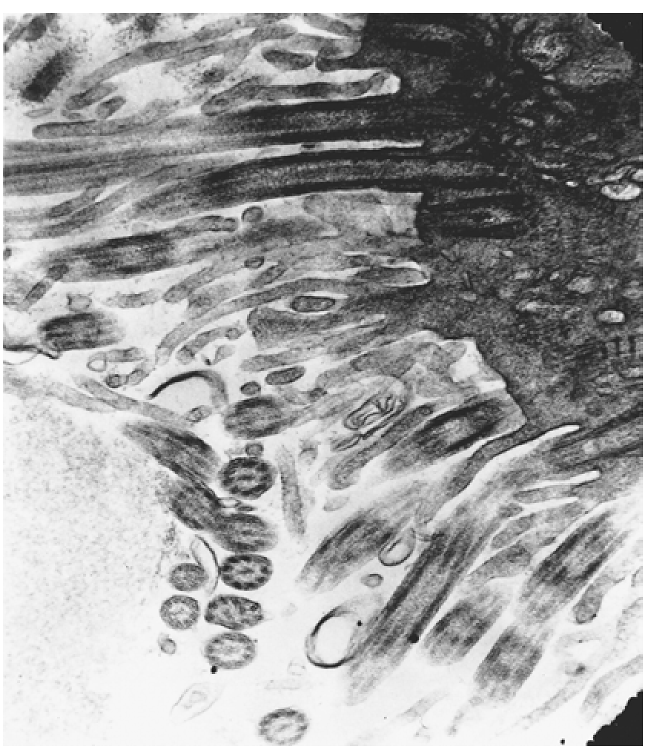


Figure 39–3 *Electron microscopy of cilia. (From Anand VJ, Pange WR. Practical Endoscopic Sinus Surgery. New York: McGraw-Hill; 1993. Reprinted with permission from the McGraw-Hill Companies.)*

Unfortunately, when any part fails, the whole cilia can fail to function properly.

Cilia beat between 10 and 15 times per second in a normal environment. They do not move as a group, but instead individually in concert. The hydrodynamics of ciliary motion have been described as "an oar moving in molasses" or a "windshield wiper." The biphasic motion has an effective phase and a slower recovery phase. When the cilia are stretched out, they sit in the gel layer, where the cells move mucus more effectively. In the recovery phase of the ciliary beat, each cilia floats in a bent fashion through the periciliary layer of mucus. The tip of the average cilia moves 0.8 cm per minute during a normal period, when it may be moving at ~1000 beats per minute.

Vascular System

There is a complex vascular system that is encountered in the nasal cavity, composed of at least five groups of blood vessel types, precapillary resistance vessels, capillaries, veins, venous erectile tissue or sinusoids, and arteriovenous anastomoses. The exact mechanisms controlling these vessels are not fully understood; however, the vascular tone of the capacitance and resistance of vessels are influenced by local metabolic and vasoactive substances, by temperature, and by neurotransmitters. In addition to the adrenergic and cholinergic neurotransmitters and receptors, several other mediators are known to be present (see **Table 39–1**). The sympathetic innervation maintains a continuous vasoconstrictor tone on the nasal mucosa. This effect is mediated by the α-adrenergic receptor system. This is reproduced or enhanced by the application of sympathomimetic drugs. The parasympathetic system mediates congestion and increased secretion. Atropine blocks the secretory effects of cholinergic stimulation. There is evidence that vasoactive intestinal peptide (VIP) may be important in regulating mucosal vasodilatation, and substance P induces vasodilatation and hypersecretion and increases the permeability of nasal mucosa.

Table 39–1 Pharmacology of Endogenous Substances on Nasal Mucosa

Vasodilators	Vasoconstrictors
Histamine	Adrenaline
Vasoactive intestinal peptide	Noradrenaline
Substance P	
Prostaglandin E_2	
Bradykinin	

Estrogen, thyroxine, cigarette smoke, ammonia vapor, and ethanol have all been shown to cause engorgement of nasal mucosa and increase nasal resistance to airflow.

Mucus

Mucus forms primarily from goblet cells with some contribution from submucous glands, transudation, serous cells, and simple ion transport. Interestingly, goblet cells are not uniformly distributed in the nose. According to Berger, goblet cell density does not appear to be influenced by the presence of perennial allergic or nonallergic rhinitis. Most of the goblet cells are located in the posterior part of the nose and nasopharynx, which is probably a major reason for the generalized condition of postnasal drip. The two most densely packed sinuses are the ethmoid cells and the maxillary antra. The mucous layer changes every 10 minutes, which is why there are such densely packed goblet cells in certain areas.

It has been estimated that the volume of nasal mucus secretions ranges from 600 to 700 mL per day. Nasal mucus is composed of 95% water, 3% mucin, and 2% other particles. Maremmani used electrophoresis to reveal the proteins present in nasal secretions, which include immunoglobulins, prealbumin, albumin, lysozyme, and several others that have yet to be identified. Mucus serves several very important functions, including moving protective particles to local infections, transporting noxious elements out of the nose, helping with humidification, and aiding in olfaction. The mixture of ingredients in mucus changes based on the condition of the nose, the general health of the person, and other local pathological problems. In the dry portion of mucus, glycoproteins play an important role. The major purpose of the glycoproteins is to trap foreign particles, which helps in their clearance. Immunoglobulins also make up a portion of nasal secretions. Immunoglobulin A (IgA) and G (IgG) are the two most commonly seen, with IgE and IgM being less common. IgA has antiviral capabilities in addition to its antiallergenic capabilities. IgE and IgG also serve as protective antibodies.

There are two types of mucus in the nose, a "sol" and a "gel" layer, which can be separated in vitro by centrifugation. The mucous glycoproteins tend to collect in the gel layer, with the serum proteins in the sol layer. In vivo, the mucus is divided into a thin periciliary phase, which is necessary for proper ciliary movement. A thicker "blanket" of mucus helps move particles by the cilia. Although the effective movements of cilia occur only at their tips, both layers are very important for proper mucociliary function. The movement of cilia is also dependent on the viscoelasticity of the mucus. Pathogenic conditions such as bronchitis and asthma may make the mucus more viscous. In addition, these two conditions may make the mucus more alkaline, which can break down cilia. In contrast, infection makes the mucus more acidic.

Cystic fibrosis is the classic mucus disorder involving hyperviscosity and abnormal sodium chloride concentrations. In general, this increase in mucus viscosity and loss of autonomic control leads to significant mucus plugging and chronic infection of the airways. Nasal polyps occur in up to 26% of patients, and over 90% have radiographic changes in the sinuses. The nasal epithelium in these patients shows hugely engorged, extruding goblet cells. The usual result is distortion of the cilia and a breakdown in the normal mucociliary clearance of the nose.

Mucus can be cleared from the nose and sinuses by several mechanisms. In the nose and sinuses, ciliary beating is a common method. In the nose, gravity, differential airflow, and sneezing are other modalities. In contrast, in the sinuses, the movement of mucus may depend on secretory pressure or the pressure gradients that may develop between the nose and the sinuses due to obstruction, sneezing, sniffing, and the simple act of respiration. The direction of mucus flow in the nose mimics the flow of air as described earlier. Anteriorly, gravity moves larger particles over 5 μm down and out of the nasal vestibule. The major flow on the lateral wall is along the middle meatus and the top of the inferior turbinate. Gravity, anatomical variations, and turbulence mediate the routes along the nasal septum.

Many factors affect efficient ciliary motion. Dehydration stops cilia, which may occur with diuretic and decongestant medications. It also may be seen in patients with deviated nasal septums, where airflow in the front of the nose desiccates the ciliated septum and anterior turbinates. The optimum temperature for ciliary motion is between 32°C and 37°C. Infections may lower the pH below the normal 5.5 to 7.0 range, which will destroy cilia. The average otolaryngologist uses many medications in the nose that may affect the cilia. Epinephrine in a 1:1000 dilution causes ciliary death, whereas a 1:10,000 dilution causes reversible inhibition. Ten percent cocaine induces paralysis, whereas a 2.5% solution only slows or stops the cilia.

Ciliary insufficiency can be caused by many acquired and congenital factors. For example, viral infections can cause clumping of the cilia. Mycoplasmal infections may cause disorganization of the cilia and ineffective motility. *Pseudomonas* overgrowth induces a detachment of the ciliated cells. Allergic reactions promote a loss of basal bodies. In addition, cigarette smoking reduces the ciliary beat frequency, and a vitamin A overdose may cause squamous metaplasia with changes in the ciliated cell concentration. Al-Rawi studied patients with severe chronic sinusitis and found that there is a variable loss of differentiated epithelial cells ranging from denuded epithelium to basal cell hyperplasia often associated with squamous metaplasia, secondary to the chronic sinonasal disease. In addition, ciliary dyskinesias may result from chronic sinusitis. The two most common congenital cilia deficiencies are Kartagener's syndrome and immotile cilia syndrome. Kartagener's syndrome presents as the triad of chronic bronchitis, situs inversus, and sinusitis. It occurs in one of every 68,000 births. The basic problem with the cilia is a lack of dynein arms, which causes a lack of energy to move the cilia. This induces the upper and lower respiratory abnormalities. Immotile cilia syndrome has a similar presentation except that it does not have situs inversus. Other congenital abnormalities include problems with the central spokes, sheath, and microtubules.

Cilia can be affected by many common factors. One of the most widespread problems is the common cold, which is caused by several different viruses. In several studies, changes were found in the ciliated cells following a cold, which remained for up to 12 weeks. This explains the chronic mucociliary problems, rhinitis, and postnasal drip that present after a cold.

Mucociliary flow in the nose is generally toward the nasopharynx. In the lungs, it is toward the throat and in the sinuses, toward the natural ostia. In the average day, the nasal cycle, dampness, cold, and age may dynamically vary the mucociliary flow. In a study in Greenland, the range of normal showed a clearance rate of 3 to 25 mm/minute. Paradoxically, 20% of the study group had no complaints but had slow clearing times.

In the maxillary sinus, Messenklinger found that the natural mucociliary flow is circular toward the natural ostia in the rabbit model. Studies from Johns Hopkins suggest that even inferior meatal punctures do not disrupt this flow pattern. Work by Hilding on the frontal sinus shows a similar pattern except in the area of the frontal duct, where there is turbulent flow. The exact nature of this clockwise flow on one side and counterclockwise flow on the other is unknown. The ethmoid sinuses are too small to have any clear understanding of their flow at this time.

Although it is widely believed that the climate affects mucociliary clearance, there are many studies that disprove this notion. Proctor and Forbes studied subjects in Denmark under different relative humidity and temperature conditions and found small changes "of little physiologic significance." Factors such as pollution, local infection, local irritants, and systemic conditions have a greater effect on nasal mucociliary clearance. Talbot has evidence to suggest that buffered hypertonic saline nasal irrigation improves mucociliary transit time.

NEUROGENIC FACTORS, NASAL CYCLE, AND NASAL RESISTANCE

Neuroanatomy

The nose is innervated by several cranial nerves as well as the autonomic nervous system. The olfactory nerve [cranial nerve (CN) I] and the trigeminal nerve (CN V) are the two primary sensory nerves, although crossover to the facial, glossopharyngeal, and vagus nerves has been described. The olfactory nerve enters the nose via the cribriform plate and innervates the upper parts of the nose. The trigeminal nerve enters the nose via the sphenopalatine foramen. The major ganglia that are involved are the gasserian ganglia and the sphenopalatine ganglia. Postganglionic fibers from the sympathetic and parasympathetic systems pass through the sphenopalatine ganglia. There is a continuous level of sympathetic tone, mediated by nerve fibers containing noradrenaline located beneath the surface epithelium and around the seromucinous glands. The sympathetic system is thought to affect nasal congestion and decongestion directly through its activity on adrenergic receptors within the erectile tissue of the nose. Parasympathetic fibers to the nose arise in the superior salivatory nucleus of the brainstem and relay in the sphenopalatine ganglion. Parasympathetic stimulation via acetylcholine primarily affects glandular secretion and may have little impact on the erectile tissue. In addition, there are numerous neurotransmitters and chemical mediators present within the nasal mucosa that may have variable effects on the erectile tissues within the nose (see **Table 39–1**).

Nasal Cycle

Historically, the nasal cycle has been described as a regular reciprocating pattern of congestion and decongestion in either side of the nose, with total airflow remaining constant. The existence of a nasal cycle has been known since first described by Kayser in 1895. It is generally accepted that a nasal cycle exists in 80% of adults. Acoustic rhinometry, which allows for the detection of very small changes in nasal volume, has confirmed the presence of a nasal cycle in most adults and in children as young as 3 years old. The presence of a nasal cycle in laryngectomy patients who have no nasal airflow suggests an autonomic control of this phenomenon. The duration of the cycle varies from 2 to 7 hours and is found throughout the day. Mizra found that older subjects were less likely to experience a nasal cycle than younger subjects. The nasal cycle can be modulated or overridden by exercise, which increases sympathetic tone, or any increase in $PaCO_2$, resulting in vasoconstriction and decreased nasal resistance. Topical decongestants also temporarily abolish the nasal cycle.

Sneeze Reflex

The sneeze reflex involves the trigeminal nerve, muscles of respiration, and the autonomic nervous system. Activation of the trigeminal nerve by thermal, chemical, or physical stimuli produces sneezing in addition to congestion and secretion. Topical nasal mucosal application of histamine induces sneezing within minutes. There is evidence to suggest that substance P may be involved in the sneeze mechanism as well.

Nasal Resistance to Airflow

Unilateral nasal resistance to airflow has been shown to be affected by changes in body position. On changing from an erect position to a recumbent one, there is an increase in total nasal resistance to airflow. Changing from the erect or supine position to the lateral recumbent position results in the congestion of the dependent nasal cavity. The application of axillary pressure results in ipsilateral nasal congestion and contralateral decongestion.

NASAL IMMUNE FUNCTION AND ALLERGIC RESPONSES

The basic principles of allergic diseases are discussed in other chapters of this text. In this section we will discuss the basics of allergy and immunology as they apply to the nose and paranasal sinuses.

The intricately complex system of mucociliary transport provides protection for the nasal mucosa; however, there are many environmental factors that can easily weaken the integrity of this barrier. In addition, genetic predisposition determines an individual's vulnerability to specific environmental substances, which is clinically manifested as the "allergic reaction," and in the nose is called "inhalant allergy." People who have allergies are often described as "atopic" individuals and have a significant nasal response to low concentrations of antigen. Nevertheless, some levels of circulating IgE are present in all individuals, and if the individual is exposed to a large enough dose of antigen, some nasal distress will be experienced. Allergic rhinitis affects ~20% of the U.S. population.

Immunoglobulins

An inhalant allergic reaction is an immune reaction to foreign material that has crossed the nasal mucosal barrier. This foreign material usually includes ordinary substances that seem to be inoffensive to the general population. The expression of an allergic reaction is dependent not only on which specific allergen is present but also on the size of the antigenic molecules and the concentration of the allergenic exposure. When antigens or allergens are introduced to the body via mucous membranes (i.e., local respiratory tract infections or inhalation of pollens), they may initiate a combined IgG/IgM response. In addition, transmucosal exposure to airborne antigens stimulates the production of IgA and IgE antibodies by the respiratory mucous membrane. This reaction causes nasal obstruction and rhinorrhea.

Many inhalant allergic problems begin in early childhood and may become progressively bothersome in early adulthood. In the young child, allergies are less common because of the small amount of IgE that is formed. IgE normally does not form until after the ninth month and increases until age 2. A second peak of serum IgE occurs in the 60s. After this age, the level may decrease, which explains why one may notice a decreasing number of allergic individuals in the geriatric group. However, because of the frequency with which people relocate to different geographic regions, allergies may often first appear in adulthood, even in the geriatric group. Frequently, rhinorrhea in the elderly is misdiagnosed as allergic, when it may be due to the slow failure of mucociliary function. Similarly, it is not uncommon for children to "grow out of their allergies"; this may be related to environmental changes, such as going off to college or moving to a new area to live.

Most individuals call atopy "hay fever," which refers to the original description of this problem in England in the late 1800s. Nevertheless, inhalant allergies may manifest themselves anytime of the year. Some allergies occur in specific seasons, such as tree pollens in the spring, grass pollens in the summer, weed pollens in the fall, and heating unit fungi in the winter. Other allergies may occur in specific locations, such as indoor dusty environments and cockroach-infested inner city apartment buildings. Some individuals respond to nonspecific airborne allergens such as nicotine in cigarettes, gaseous breakdown products from automobile combustion, or the chemicals from cleaning agents. Uniform to all these individuals is usually some nasal reaction. Other forms of the allergic response may include asthma, dermatitis, or gastric distress.

IgA is considered to be the first line of defense for allergies because it is present in secretions such as nasal mucus and the tears that run through the nose. These antibodies are produced locally in patches of

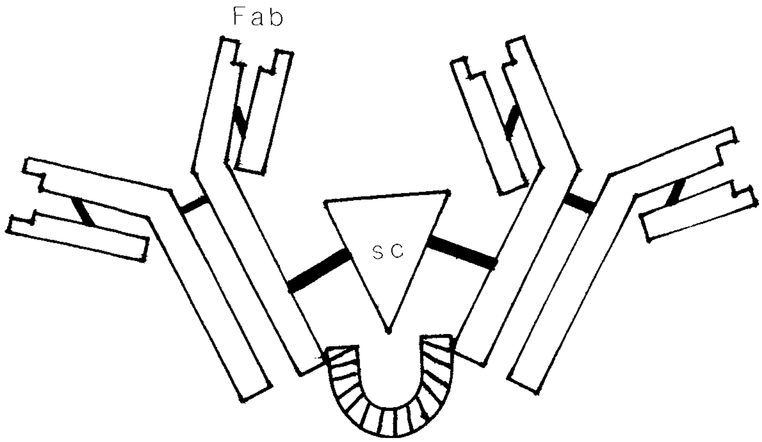


Figure 39–4 *Secretory immunoglobulin A (SIgA). Secretory component (SC) is produced by serous-type epithelial cells near the IgA immunocytes. SC combines with dimeric IgA at the basolateral plasma membrane, and SIgA is then discharged into the mucous layer. Note the J-chain that joins two monomeric IgA molecules at the Fc region. SIgA has four antigen-binding sites. (From Anand VJ, Pange WR. Practical Endoscopic Sinus Surgery. New York: McGraw-Hill; 1993. Reprinted with permission from the McGraw-Hill Companies.)*

mucosa-associated lymphoid tissue (MALT) that are located on the inferior and middle turbinates of the nose. IgA is a monomeric antibody that is produced by specialized immunocytes that are also capable of producing J-chains. When two monomeric IgA molecules are united by a J-chain, the resulting dimeric structure has a binding site for secretory components (**Fig. 39–4**). The secretory component is produced in serous-type secretory cells (i.e., nasal acinar and duct cells) and migrates to the basolateral plasma membrane, where it combines with the dimeric IgA molecule. Thus secretory IgA (SIgA) complexes are formed at the plasma membrane and are later "packaged" into vesicles by the process of adsorptive endocytosis. These vesicles can transport the SIgA across the cell membrane and release them into the mucous layer in high concentrations. Each molecule of SIgA has four binding sites for antigens. The major function of SIgA in mucous secretions is to bind to airborne antigens and to prevent them from adhering to the epithelial surface of the mucous membrane, which prevents transmucosal invasion.

In addition to its effect on aeroallergens (dusts, pollens), secretory IgA participates in the opsonization of airborne bacteria, which maximizes IgG-mediated phagocytosis. IgA also has been found to have bactericidal effects when functioning with lysozyme and complement. Aggregated IgA complexes are known to activate the classic complement pathway.

The most common types of immune deficiency conditions known are IgA deficiencies. Although some people who have low titers of IgA are apparently healthy and free of symptoms, many people with IgA deficiencies suffer from multiple allergies and sinopulmonary problems. It therefore seems that when the first line of immune defense is inadequate, airborne antigens can more easily cross the nasal mucous membrane barrier.

The second line of defense is IgG, which helps to form immune complexes. Like IgA, IgG is produced in mucus-associated lymphoid tissue in large concentrations, but there is very little uptake in the glandular acini as there is with IgA. Therefore, less IgG is transported to the mucous layer in this manner. However, monomeric IgG can diffuse passively through capillaries and subsequently appears in nasal secretions. Increased concentrations of IgG are commonly found during acute respiratory infections because the inflamed and damaged mucous membrane is even more permeable. IgG-mediated phagocytosis of organisms and its potential cytotoxic effects suggest that IgG plays an important role in immune exclusion.

The most important immunoglobulin for the allergic response of the nose is the IgE particle. IgE mediates the type I of hypersensitivity reactions, which have been described by Gell and Coombs. Type I is an immediate hypersensitivity reaction in which IgE-sensitized cells immediately degranulate upon contact with specific antigens. However, this does not occur during the first exposure to the antigen. The first time a specific antigen penetrates the mucosal barrier, stromal macrophages and dendritic epithelial cells (i.e., Langerhans' cells) will bind the antigen and subsequently "present" it to T-helper lymphocytes. T-helper cells then mediate the subsequent proliferation and differentiation of B lymphocytes via interleukins (particularly IL-4 and IL-13) into specific IgE-producing immunocytes.

IgE antibodies have a cytophilic affinity for the plasma membrane of tissue mast cells and blood neutrophils, basophils, and eosinophils. This trait is regulated by the Fc region of the epsilon heavy chain (**Fig. 39–5**). When IgE antibodies become attached to these cells, the cells are then "sensitized." When sensitized cells are later reexposed to the same allergens, they immediately rupture upon contact and undergo degranulation, releasing various vasoactive mediator substances and

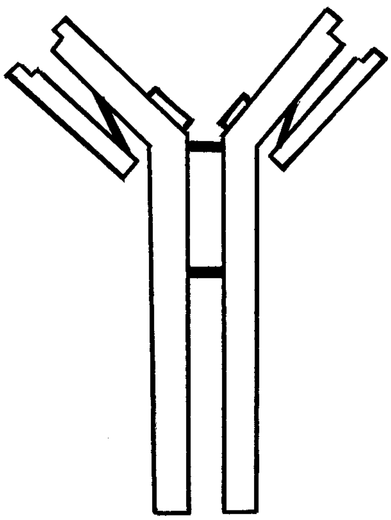


Figure 39–5 *IgE Fc region. Affinity for mast cell binding is at the Fc-receptor site. (From Anand VJ, Pange WR. Practical Endoscopic Sinus Surgery. New York: McGraw-Hill; 1993. Reprinted with permission from the McGraw-Hill Companies.)*

triggering the production of others. These substances are potentially toxic to local tissues and are responsible for the clinical symptoms of allergy that develop (i.e., rhinorrhea). There is an immediate reaction to allergen exposure as well as a late-phase response that may be seen 3 to 12 hours after the exposure. The late symptomatology includes nasal congestion and reflects inflammatory mediator production and cell influx, particularly eosinophils and T-helper cells.

Immune Cells

Promyelocytes are derived from pluripotential stem cells in bone marrow and possibly in thymus. These cells mature into granulocytes that can be found circulating in the blood, as well as in the tissues where they accumulate during acute local infections. In the blood, they constitute ~60 to 75% of all white blood cells. During infectious or inflammatory conditions they circulate in even higher numbers. They represent the body's mobile "armed forces," protecting it from all types of infection and inflammation. Several types of granulocytes exist (**Fig. 39–6**).

The eosinophils are not the most prominent subgroup but are extremely important to parasitic and allergic types of reactions. Although only 2 to 5% of white blood cells are eosinophils; higher concentrations are found in allergic states. These cells bear many receptor sites for specific IgE molecules on their surfaces, and once attracted by eosinophil chemotactic factor, they become a formidable force against invading allergens. Upon contact with such an antigen, they degranulate and release extremely toxic substances.

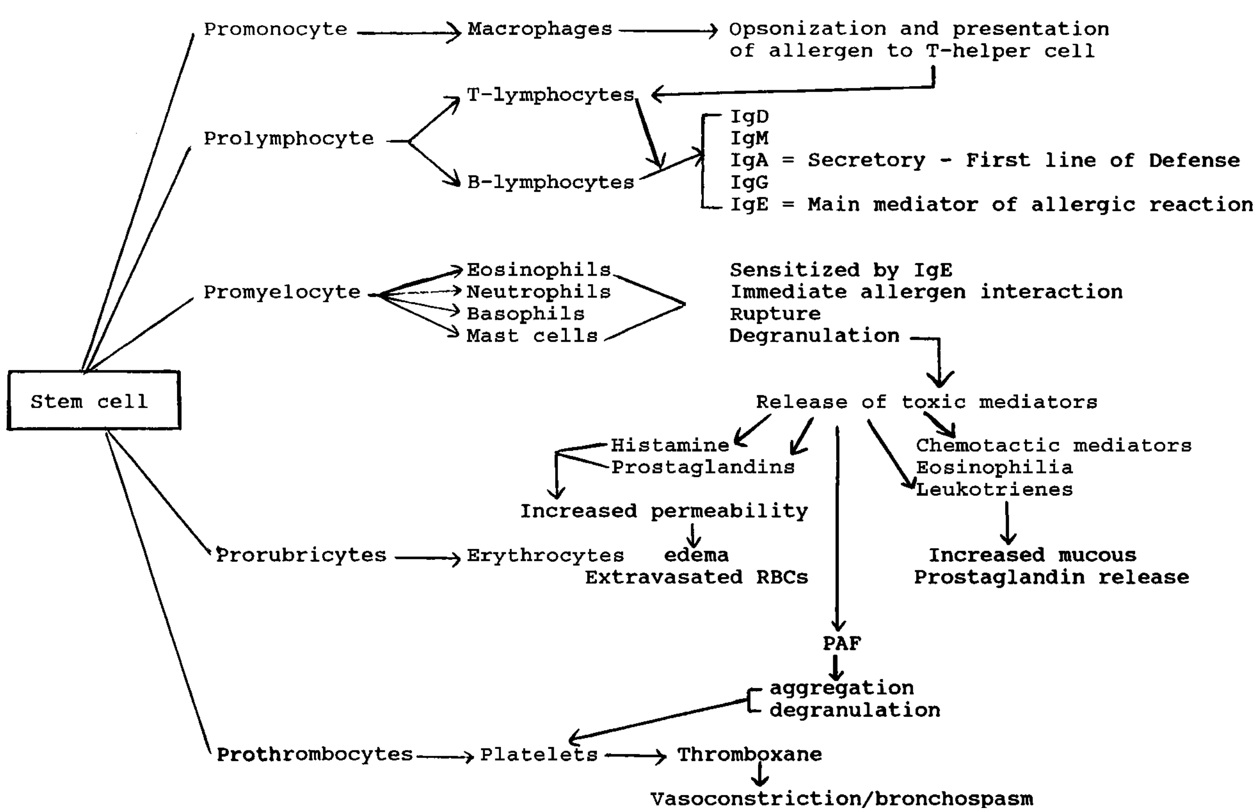


Figure 39–6 *Schematic diagram of the cells, interactions, and toxic mediators involved in an allergic reaction. (From Anand VJ, Pange WR. Practical Endoscopic Sinus Surgery. New York: McGraw-Hill; 1993. Reprinted with permission from the McGraw-Hill Companies.)*

Basophils are another subgroup of circulating granulocytes that can be activated by binding with IgE. Tissue mast cells arise in a similar manner from promyelocytes that come from totipotential stem cells in the bone marrow. They are often compared with basophils and are sometimes referred to as "tissue basophils." There are two types of these cells, including connective tissue and mucosal mast cells. The proliferation of the mucosal mast cells is dependent upon T-cell mediation. Mast cells and basophils bear many thousands (50,000–500,000) of receptor sites for IgE. When two IgE molecules adhere to one of these cells, the phenomenon of bridging takes place via the Fc regions. Bridging provides a strong activating signal and enhances their highly sensitive and specific surveillance function. When specific allergens are then encountered, the reaction that follows results in swelling of the cell, increased intracellular metabolism, and the production of arachidonic acid. Eventually, the cell expels many granules of histamine, as well as the products of arachidonic acid metabolism (leukotrienes and prostaglandins).

Thrombocytes (platelets) are derived from another progenitor cell line (prothrombocytes), but basically they originate from the same type of pluripotential bone marrow stem cell. During acute allergic reactions, these are also activated by IgE binding, which causes them to aggregate and rupture. The results of these actions may lead to acute, severe asthmatic bronchospasm and anaphylaxis.

Immune Mediators

There are several vasomotor mediator substances that are known to be released or activated by IgE-triggered immune mechanisms, especially in the nose. These include histamine, platelet-activating factor, leukotrienes (slow-reacting substance of anaphylaxis), eosinophil chemotactic factor, major basic protein, and prostaglandins. These mediator chemicals can produce major changes in the lining of the nose and sinuses and cause itching, watery discharge, a change in the mucociliary flow, and disruption to the olfactory areas.

Histamine is the most well known of the mediator substances. It is stored in the lysozymes of tissue mast cells and has many known effects on various systems. First, it has potent vascular effects, such as increasing the permeability of capillaries, leading to leakage of fluid into the extravascular space. This causes congestion and swelling of nasal mucosa and increased mucus production, and may also lead to acutely depressed blood pressure associated with dizziness, syncope, and cardiovascular collapse. Second, histamine may have respiratory effects, by causing bronchiolar smooth muscle to contract, leading to bronchospasm. Third, it may have a vagal influence by stimulating exocrine glands, thereby producing bronchial hypersecretion (as in asthma) and nasal hypersecretion (as in hay fever and rhinitis). Fourth, it may have a powerful immunologic effect. It does this by suppressing T-lymphocyte proliferation and inducing monocytes to produce prostaglandins (this effect is usually seen only in atopic individuals). Last, histamine may have a hematologic effect by mediating the activation of platelets in the nose, promoting slight nasal epistaxis.

A second common mediator substance is the platelet-activating factor, which is produced by mast cells, basophils, eosinophils, and neutrophils in response to IgE-mediated activation. It has an effect on platelet and neutrophil aggregation and may also alter vascular permeability. This factor has been associated with shock because it causes vasoconstriction and cardiovascular collapse. It may also lead to cardiac arrhythmias, pulmonary artery hypertension, and pulmonary edema.

A third mediator is prostaglandin, which is produced from the metabolism of arachidonic acid derived from stored phospholipids and triglycerides in mast cells. The biosynthesis of prostaglandins involves complex enzymatic reactions that may be triggered by IgE-mediated allergic reactions. Prostaglandins GD_2, GF_2, and GE_2 cause bronchiolar smooth muscle contraction, prevent platelet aggregation, and have a chemotacticeaffect on leukocytes, markedly potentiating the inflammatory effects on local tissues.

Leukotrienes are another group of extremely toxic mediators that are derived from stored arachidonic acid in mast cells. They are thought to be even more potent in their effects than histamine. Leukotrienes C, D, and E can also generate prostaglandins and increase mucus production. Leukotriene C (previously called "slow-reacting substance of anaphylaxis") causes a slow but prolonged contraction of venous smooth muscles and can increase vascular permeability, causing increased tissue swelling and vasomotor collapse. Leukotriene B has a chemotactic function and also stimulates the release of histamine from mast cells and neutrophils.

Other vasoactive mediators include adenosine, eosinophil chemotactic factor, eosinophil cationic protein, eosinophil peroxidase, and thromboxane. Adenosine is derived from the breakdown of adenosine triphosphate (ATP) and is released from tissue mast cells during IgE-mediated reactions. Although it has no effect on bronchiolar smooth muscle in nonatopic people, it causes bronchospasm in asthmatics. It also has strong vasodilator effects and causes tissue edema, especially in the nose. Eosinophil chemotactic factor is released from mast cell granules and is responsible for attracting great numbers of eosinophils into the allergically inflamed area. Like mast cells, eosinophils can be "activated" by IgE-mediated interactions with allergens and, upon contact, release other toxic mediators that are stored in granules and enhance the inflammatory reaction. Eosinophil cationic protein and eosinophil peroxidase are stored in granules and are extremely toxic when released in the nose. The peroxidase can participate in creating a cytotoxic environment for mast cells, bacteria, and parasites. Thus one of its effects is mast cell degranulation, with subsequent further release of toxic mediator substances. Thromboxane is another mediator and is released during the degranulation of platelets and macrophages. It is a potent vasoconstrictor and causes severe bronchospasm.

These are just a few of the chemical substances that are produced when an allergen encounters various cell mediators of immunity. It is apparent that their prompt and potent effects not only are effective in immune elimination of foreign antigen but also, unfortunately, are toxic to the body's own tissue. At the level of the nasal mucosa, this results in the manifestations of rhinitis.

PHYSIOLOGICAL TESTING

Physical Examination

The primary modality for viewing the nose is a thorough physical examination. This entails inspecting the nose with a nasal speculum before and after phenylephrine spray. An initial examination of the nasopharynx with a small mirror is also important.

Nasal Endoscopy

After spraying with a topical anesthetic, a nasal endoscope can be used to look at the inferior and middle meatuses. If possible, the area of the nasolacrimal duct and the natural ostium of the maxillary sinus should be inspected. Posteriorly, the sphenoid ostia, fossa of Rosenmüller, eustachian tube, and nasal portion of the soft palate can be examined.

Nasal Cytology/Nasal Irrigation

There are several possible ways to test the nasal mucus and cells without being excessively invasive or disturbing to the patient. First, nasal cytology can help to identify the quantity of nasal eosinophils and mast cells in allergic reactions or lymphocytes and neutrophils in infection. This can easily be performed using a microswab and the Wright-Giemsa stain. Nasal irrigation is another method used to sample epithelial lining fluid for inflammatory mediators and cells produced during nasal mucosal inflammation. Eosinophil count from nasal irrigation has been shown to be a sensitive marker of perennial allergic rhinitis. Second, the pH of the mucus can be determined with a simple meter. This will help to identify infectious states where the pH may be diminished. In acidic environments, the cilia work poorly, and patients may demonstrate abnormal flow patterns. Third, the viscosity or elasticity of mucus can be studied by using a viscometer or rheometer (galvanometer). The relative proportions of glycoproteins and ion content may affect the viscosity and elasticity of nasal mucus that will ultimately change the mucociliary flow.

Rhinomanometry

One of the easiest ways to determine the airflow patterns and patency of the nasal cavity is to perform rhinomanometry. This test gives an objective reading of flow versus differential pressure. Franke in 1894 first devised a model for measuring the airflow through a system looking like a nose. He called this anterior rhinometry. Kayser, in 1895, modified this technique by blocking off one side of the nose with an olive to measure pharyngeal pressures. This was called posterior rhinometry. Since then, several models have been transformed into the computerized tests that are now available. In these new pieces of equipment, either nasal cannulas or face masks can be employed.

The basic physics of nasal airflow are vital to our understanding of this test. Ohm's law states that resistance (R) equals pressure (P) divided by flow volume (V) ($R = P/V$). This presumes normal laminar flow. When turbulent flow is encountered, the formula changes to $R = P/V^2$. The real figure in the nose is not 2 but 1.7. In addition, the nasal airflow is directly related to the cross-sectional area of the radius of the nose to the fourth power. This means that increasing the inner diameter of the nasal cavity can increase the airflow dramatically.

The anterior rhinomanometer displays a flow volume loop similar to pulmonary function tests (**Fig. 39–7**). The flow is measured in cm^3/second and the pressure in pascals. The expired air is to the left and inspired to the right. Whenever there is obstruction, the flow curve will be depressed. When performing this test, it is important that the patient is breathing normally and that the instruments, masks, or nasal cannulas do not distort the nasal passages, because this could significantly distort the results.

Acoustic Rhinometry

Acoustic rhinometry evaluates the geometry of the nasal cavity with acoustic reflections and provides information about nasal cross-sectional areas and nasal volume. It is a noninvasive test that is easily performed

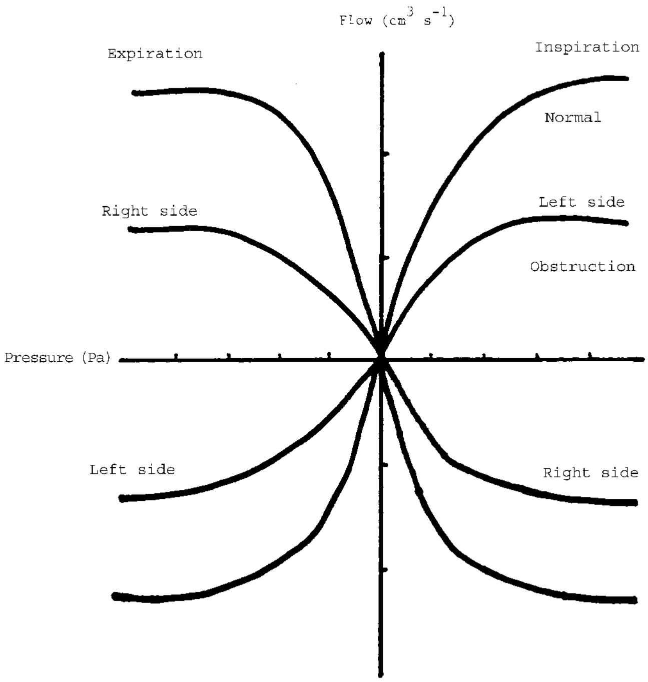


Figure 39–7 *Flow volume loop: anterior rhinometer. (From Anand VJ, Pange WR. Practical Endoscopic Sinus Surgery. New York: McGraw-Hill; 1993. Reprinted with permission from the McGraw-Hill Companies.)*

in a short period of time with minimal patient cooperation needed; thus it may be used to evaluate infants and children. A sound pulse is obtained from the sound generator, which propagates through a tube and enters the nasal cavity through a nosepiece. The incident and reflected signals are then measured by a microphone in the sound tube to calculate corresponding cross-sectional areas and volumes. Acoustic rhinometry has been used as an objective measure of nasal patency after nasal septal and turbinate surgery and has been shown to correlate with subjective measures of airflow.

Tests of Ciliary Structure and Function

Effective mucociliary clearance depends on a normal number of cilia per cell, a normal rate of beating, and coordinated motion. There are several ways to measure these parameters. The two most commonly used methods are the use of light fluctuations induced by cilia, which can be monitored and measured by a photomultiplier, and ultra-high-speed photography. In the first method, ciliated cells are scraped from the nose and placed on a slide under a phase-contrast microscope and viewed with a special fiberoptic probe by a photomultiplier. The results can be plotted on a strip of paper or in a computer for analysis. In the second method, the harvested cells are viewed using high-speed photography at 400 frames per second. This gives a good (although very expensive) picture of the ciliary action. Electron microscopy may be performed on a nasal biopsy specimen to evaluate the nasal cilia structure.

Ciliary beat frequency tests the viability and integrity of these cells in in vitro form. In a patient with bronchiectasis, the movement of the pulmonary cilia is markedly slow, which results in abnormal clearance time. Alternatively, a patient with cystic fibrosis who has normal cilia but abnormal mucus will have normal ciliary beat frequency but abnormal clearance due to the deficient mucous layers. Patients with sinusitis will have abnormal mucociliary clearance that may be due to a combination of abnormal ciliary beat frequency and mucus factors.

Rhinostereometry

Rhinostereometry is an optical, direct, noninvasive method for measuring nasal mucosal swelling with a high degree of accuracy. A surgical microscope is placed on a micrometer table fixed to a frame. The microscope can be angled in three directions, establishing a three-dimensional coordinate system. The subject is fixed to the apparatus by an individually made plastic tooth splint. The nasal cavity is viewed through an eyepiece that has a horizontal millimeter scale. Because the microscope has a small depth of focus, changes in the position of the mucosal surface on the medial side of the head of the inferior concha are registered in the plane of focus along the millimeter scale. The accuracy of the method is reported as 0.2 mm. This technique is being used to study mucosal responsiveness to medications and variability of mucosal responsiveness in different disease states.

Contact Nasal Endoscopy

Another exciting tool in the early experimental phases of development that may provide great advances in the understanding and management of nasal and paranasal sinus disorders is contact nasal endoscopy. Contact endoscopy allows direct in vivo, in situ visualization of the superficial cell layers of nasal epithelium. This technique has demonstrated squamous epithelium, ciliated epithelium, glandular ostia, mucus, submucosal vascular networks, inflammatory cell infiltrates, tissue inclusions, nuclear characteristics, and fungal hyphae.

SUMMARY

An understanding of the physiology of the nose and paranasal sinuses is crucial to the diagnosis and treatment of the various diseases that may affect this organ system and the general health of our patients. Over the past century a growing fund of knowledge has evolved, allowing us to more effectively treat sinonasal disease. Despite the advances that have been made, millions of people continue to suffer from nasal and paranasal sinus disorders. Only through further study and investigation can we hope to help our patients.

Suggested Readings

Brandtzaeg P. The immune system of the nose and nasopharynx. In: Mygind N, Pipkorn U, eds. Allergic and Vasomotor Rhinitis: Pathophysiological Aspects. Copenhagen: Munksgard; 1987:98

Cole P. The Respiratory Role of the Upper Airways. St. Louis: Mosby Year Book; 1993

Greenberger PA. Nasal receptors and reflexes. Allergy Asthma Proc 1998;19(4):175–179

Lund VJ. Nasal physiology: neurochemical receptors, nasal cycle, and ciliary action. Allergy Asthma Proc 1996;17:179–184

Naclerio R, Solomon W. Rhinitis and inhalant allergens. JAMA 1997;287(22):1842–1848

Proctor DF. The mucociliary system. In: Proctor DF, Andersen I, eds. The Nose: Upper Airway Physiology and the Atmospheric Environment. Amsterdam: Elsevier; 1982:263

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. The nasal valve is composed of the following structures except the
   A. Nasal septum
   B. Inferior turbinate
   C. Lower lateral cartilage
   D. Middle turbinate
2. The smallest-size particle that can be filtered by the nose is
   A. 100 μm
   B. 10 μm
   C. 5 μm
   D. 1 μm
   E. 0.5 μm
3. Kartagener's syndrome is clinically manifested by all of the following except
   A. Situs inversus
   B. Bronchiectasis
   C. Sinusitis
   D. Cystic fibrosis
4. Which of the following can induce cilia dysfunction or cilia death?
   A. Rhinovirus
   B. 10% cocaine
   C. Epinephrine 1:1000
   D. Cigarette smoke
   E. All of the above
5. The following is true of secretory immunoglobulin A (SIgA) except
   A. Each molecule of SIgA has four binding sites for antigens
   B. SIgA is found in nasal mucus
   C. SIgA participates in the opsonization of bacteria
   D. SIgA mediates the immediate hypersensitivity reaction

Chapter 40

THE BIOLOGY AND TESTING OF OLFACTORY DYSFUNCTION

James E. Schwob, Daniel B. Kurtz, and Bradley J. Goldstein

The relevance of olfactory dysfunction is often overlooked by the clinician. However, olfaction is a special sense that can be critical to the health and well-being of the patient. The sense of smell is responsible for most of the perception of the flavor of food and is, therefore, a critical determinant of nutritional intake. The sense of smell acts as a personal warning system, alerting an individual to smoke, natural gas, spoiled food, and the presence of potentially hazardous chemicals. In addition, the sense of smell provides the patient with feedback concerning personal hygiene and adds to the quality of life in such areas as eating, connection with nature, and intimate relationships. Finally, the loss of olfaction can be a harbinger of more serious disorders such as Alzheimer's disease, human immunodeficiency virus (HIV), dementia, and Kallmann's syndrome, to name a few. The careful clinical evaluation of olfactory dysfunction provides an insight into a disorder that may carry no external manifestations but one that can inexorably alter a patient's life.

STRUCTURE AND FUNCTION

Before discussing the nature and treatment of clinical olfactory disorders, it is necessary to review the structure, normal function, and response to injury of this relatively unappreciated sensory system.

Anatomy

The lining of the nasal cavity in humans is of three types. The vestibule is lined by stratified squamous epithelium. Much of the rest is lined by respiratory epithelium. The posterodorsal recess situated immediately inferior to the cribriform plate is lined by olfactory epithelium (**Fig. 40–1**), where the olfactory epithelium originally

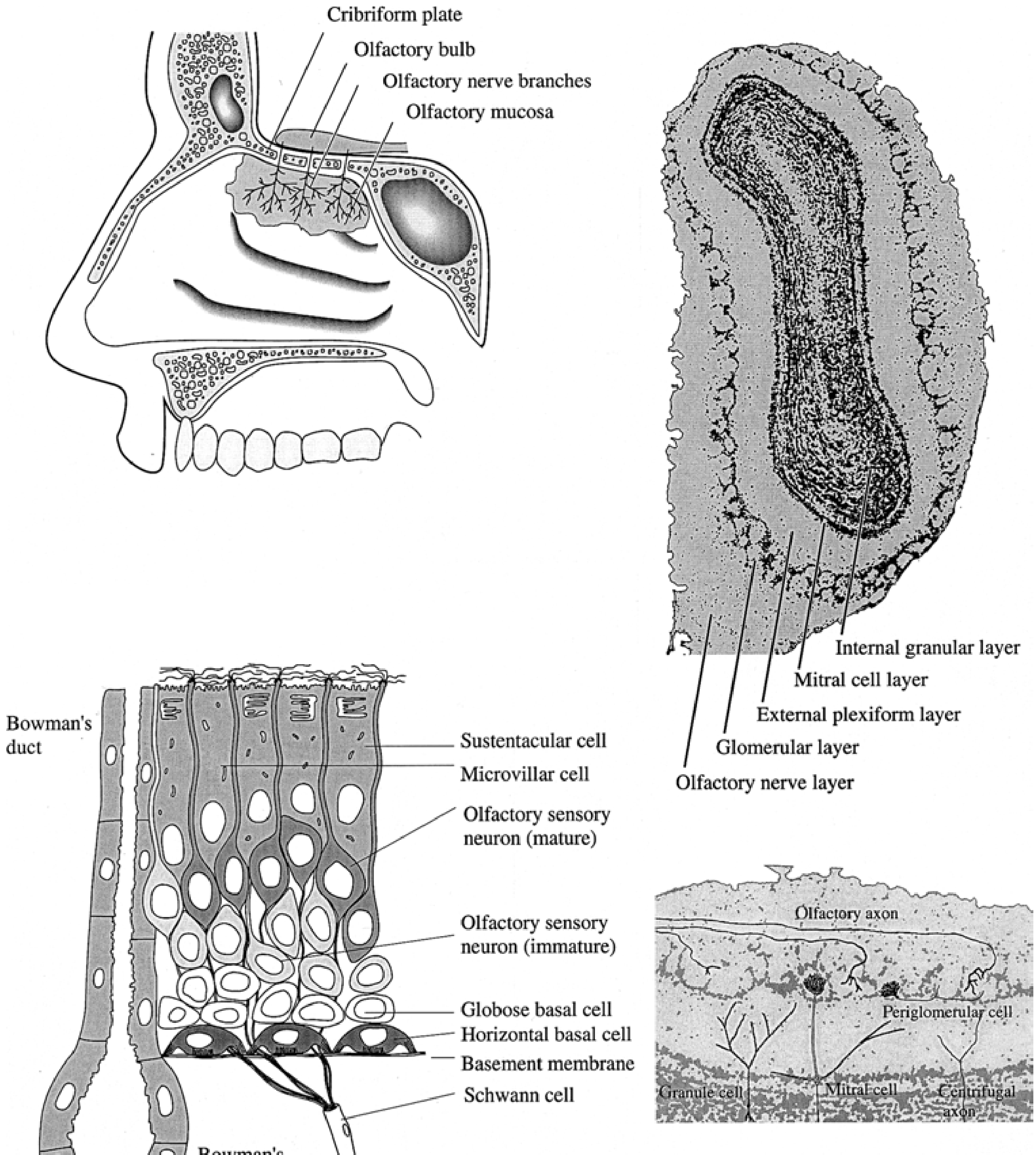


Figure 40–1 *Schematics of peripheral olfactory apparatus and the olfactory bulb. In the rendition of the lateral nasal wall, the olfactory area is shaded. However, even in normosmic individuals, substantial parts of the epithelium have undergone respiratory metaplasia. The olfactory epithelium is presented as it appears in newborns. The regular lamination of the various cell types is degraded somewhat, and the numbers of neurons are reduced as a consequence of the repeated cycles of damage and reconstitution to which the olfactory epithelium is subject during life. The olfactory bulb is that of a rodent. However, the structure and cellular constituents are closely similar in the human olfactory bulb.*

forms as a continuous sheet during embryological development. In adults, the olfactory area measures ~1 cm^2. Odorants are transported to the olfactory area with quiet breathing, but sniffing accentuates the delivery of the stimuli. The passage from the vestibule to the olfactory area is formed by the nasal septum and the inferior, middle, and superior turbinates. This passage is narrow and easily compromised by a variety of deformities and/or ongoing processes such as inflammation and infection. The nasal mucosa is also innervated by branches of the ophthalmic and maxillary branches of the trigeminal nerve.

Situated immediately above the cribriform plate are the olfactory bulbs. The bulb is the first-order relay point for olfactory input. The bulbs are tethered tightly within the olfactory grooves by the fascicles of the olfactory nerve as they traverse the cribriform plate and merge with the surface of the bulb.

Peripheral Olfactory Apparatus

The olfactory epithelium is a laminated, pseudostratified epithelium composed of a limited number of cell types (**Fig. 40–1**). These include (1) sustentacular or supporting cells, which are flask-shaped (their nuclei are the most superficial in the epithelium), (2) olfactory neurons, which are bipolar in shape (their nuclei form a layer two to eight cells thick in the middle zone of the epithelium); and (3) basal cells, which are situated deep in the epithelium in juxtaposition to the basal lamina. Two types of basal cells can be distinguished morphologically. Horizontal basal cells resemble basal cells in other epithelia; they assemble keratin-containing intermediate filaments that attach to hemidesmosomes and mediate attachment to the basal lamina. Globose basal cells are unique to the olfactory epithelium; they are relatively poorly characterized but proliferate at a much higher rate than any other cell type in the olfactory mucosa. The lamina propria, deep to the basal lamina, contains fascicles of the olfactory nerve and the specialized mucus-secreting cells of Bowman's glands; the ducts of these glands extend to the surface of the olfactory epithelium. Transduction of odorant stimuli requires the sorption of the stimuli into the mucous layer and diffusion (possibly assisted by odorant-binding proteins) to the fine, tapering immotile cilia elaborated by the apical processes of the sensory neurons. Unlike the sensory epithelia of the auditory, vestibular, or gustatory systems, the receptor cell of the olfactory epithelium is a bona fide neuron. The olfactory sensory neurons elaborate an unbranched, unmyelinated, very thin axon that may exceed 10 mm in length; it generates and conducts typical action potentials to the central nervous system (CNS). The axons of the olfactory sensory neurons join together to form progressively larger bundles within the lamina propria that traverse the cribriform plate to reach the olfactory bulb.

Central Olfactory Apparatus

The olfactory bulb has a highly stereotyped laminar organization, from outermost to deep (**Fig. 40–1**):

- The olfactory nerve layer, consisting of glial cells and the bundled axons from the nose
- The glomerular layer, composed of more than 1000 glomeruli. Each glomerulus is a pocket of neuropil surrounded by small inhibitory interneurons called periglomerular or juxtaglomerular cells. Axons from the periphery make synaptic contact with second-order relay neurons called mitral and tufted cells and with the periglomerular cells in the glomerular neuropil. In addition, the periglomerular cells and relay neurons make reciprocal dendrodendritic synapses that conduct excitatory input from the relay neurons to the interneurons and inhibitory feedback from interneurons back to mitral and tufted cells
- The external plexiform layer, a relatively acellular band of neuropil. The sensory signals are modified in this layer via lateral, feed-forward, and feedback inhibitory processing between granule cells, which are another type of inhibitory interneuron, and the relay neurons. Reciprocal dendrodendritic synapses between relay neurons and granule cells are the anatomical substrate for these physiological processes
- The mitral cell layer, composed of the major population of second-order relay neurons. The axons of these mitral cells are the major contributors to the lateral olfactory tract
- The internal granular layer, containing the somata of the inhibitory granule cell interneurons

The olfactory peduncle encompasses both the anterior olfactory nucleus and the lateral olfactory tract. The central olfactory cortical areas are those portions of the cortex in the vicinity of the limen insula and temporal pole that receive either a direct or an indirect input from the olfactory bulb.

Most olfactory symptomatology reflects disordered processes in the peripheral olfactory system. Thus the peripheral olfactory system, composed of olfactory epithelium, olfactory nerve, and first-order synaptic regions of the olfactory bulb, will be the focus of the remainder of the review.

Olfactory Sensory Neuron Function

One of the most important advances in our understanding of olfactory function and dysfunction was the identification of a very large family of genes, 500 or more in number, that encode seven transmembrane domain proteins and are selectively expressed in the olfactory epithelium. The members of this gene family function as odorant receptors. Each sensory neuron apparently expresses only one allele of one gene locus of the odorant receptor family. Most if not all receptors are broadly tuned and are activated by a broad range of compounds. Conversely, each odorant will bind to a large number of odorant receptors (i.e., activate many different functional types of olfactory sensory neurons). Sensory transduction couples the receptor to the activation of olfactory-specific adenylyl cyclase via the mobilization of G proteins (**Fig. 40–2**) increases cyclic adenosine monophosphate (cAMP) in the cilia and proximal dendrite activated cAMP-dependent ion channels in turn. The transduction cascade can be downregulated following persistent stimulation via phosphorylation of the receptor by a member of the serine-threonine kinase family. Suppression of the signal at the level of the odorant receptor may play a prominent role in the behavioral habituation to prolonged odorant exposure.

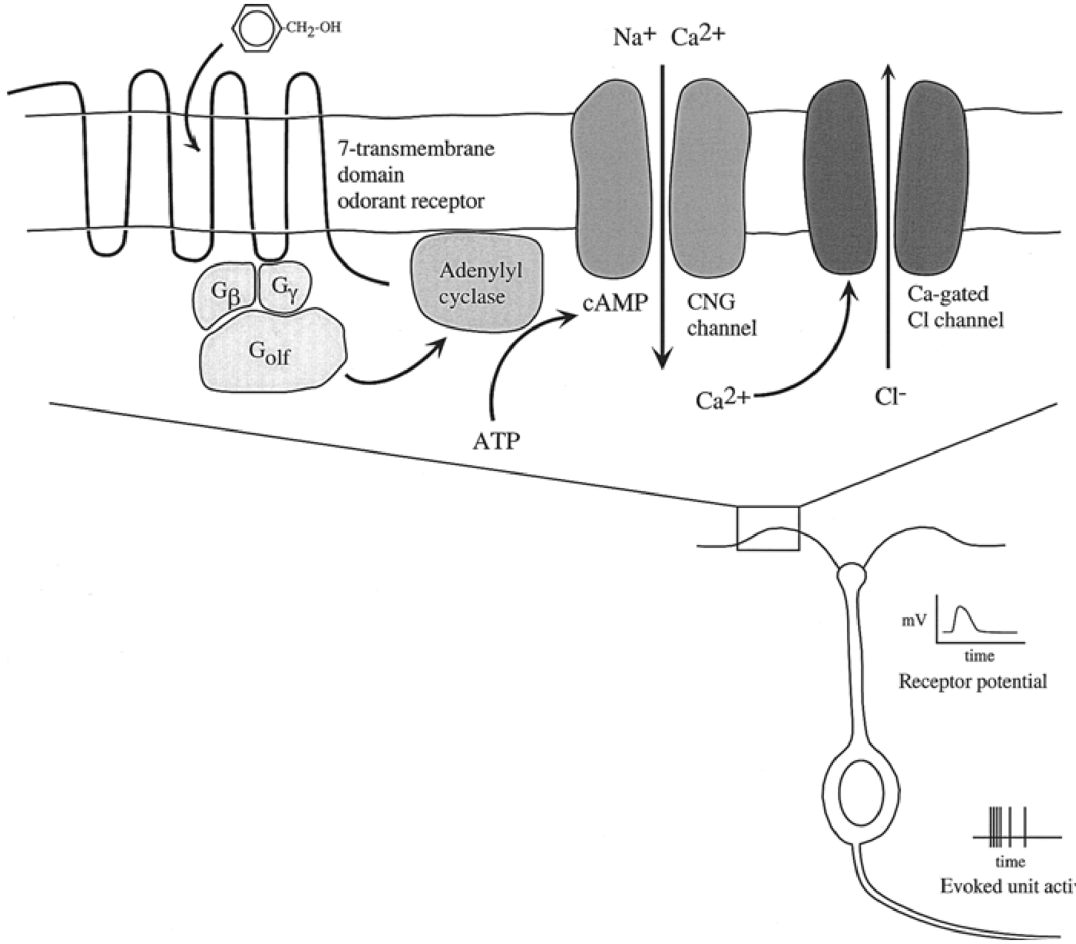


Figure 40–2 *Schematic of odorant signal transduction by olfactory sensory neurons. The events of signal transduction occur in the immotile cilia that extend from the olfactory knob and the knob itself. Transduction is initiated by interaction of the odorant with the odorant receptor; variability in protein sequence suggests that the pocket formed by transmembrane domains 3, 4, and 5 forms the interactive surface. Olfactory neurons apparently express only one member of the large family of odorant receptor genes and only one allele at that locus. A variety of studies indicate that receptors are broadly tuned and respond to a large group of odorants. Our understanding of the molecular receptive range of individual receptor types (i.e., analogous to the stimulus response profile of a receptor for peptide, hormone, or neurotransmitter) is still in its infancy. Transduction requires the activation of the trimeric G protein, of which the protein* G_{olf} *is the olfactory-specific* α*- (or stimulatory) subunit. Downstream events include the sequential activation of the olfactory-specific adenylyl cyclase by guanosine triphosphate–binding* G_{olf} *and the opening of the cyclic nucleotide-gated (CNG) nonspecific cation channel (also olfactory-specific) by cyclic adenosine monophosphate (cAMP). Gene knockout experiments that eliminate the CNG channel suggest that cAMP-mediated activation of the channel is the necessary and sufficient effector cascade for stimulus transduction. An unusual feature of the receptor potential generation by the olfactory neurons is the inward current generated by the outward flow of* Cl^- *through the* Ca^{2+}*-activated anion channel, which follows the high intracellular concentration of* Cl^- *in olfactory neurons. Downregulation of the transduction cascade is accomplished via receptor kinases homologous to the* β*-adrenergic receptor kinase (BARK) that act by phosphorylating odorant receptors (not shown). ATP, adenosine triphosphate.*

The pattern of odorant gene expression is critical not only to a sensory neuron's response spectrum but also to its axonal projection onto the bulb. The neurons that express a particular odorant receptor protein are distributed across large swaths of the olfactory epithelium. Despite the dispersion of "like neurons" across the epithelial sheet, their axons eventually gather together in the nerve layer and project onto a single glomerulus on the medial surface of the bulb and a single glomerulus on its lateral surface, a spatial convergence of roughly 3000-fold. The manner in which funneled spatially distributed neurons are onto two target glomeruli depends on the type of receptor that is expressed and is still poorly understood. The exquisitely precise and stereotyped connectivity of neurons expressing the same odorant receptor has the effect of converting a chemical stimulus into patterned neural activity across the glomerular sheet. As a further consequence, each odorant will elicit its own distinctive pattern of activity across the neural space of the bulb.

In addition to their activation of sensory neurons, some chemicals have the capacity to stimulate the free nerve endings of the trigeminal nerve, which ascends from the lamina propria through the olfactory epithelium toward its apical surface. The activation of trigeminal fibers contributes to the unitary percept associated with an odorant and conveys the quality of pungency. Thus chemicals like acetic acid (vinegar) and ammonia are particularly strong activators of the trigeminal nerve. Indeed, carbon dioxide is capable of stimulating only the trigeminal fibers and not sensory neurons, which are responsible for our ability to detect high concentrations of the gas by the "sting" it produces.

A SENSORY SYSTEM AT RISK

By virtue of the structure of the peripheral apparatus, olfactory sensory neurons are unprotected; that is, they are in direct contact with their environment, as is needed for proper chemosensory function to occur. Because they are in contact with the environment, olfactory neurons are highly susceptible to being damaged and then dying. In addition, the peripheral olfactory epithelium is only a few synapses away from the temporal cortex, and compounds or viruses from the nose can quickly reach the cerebral cortex via anterograde axoplasmic and transsynaptic transport; this route may be an important means of access for viruses such as herpes simplex and other potential pathogens.

Given the critical nature of olfactory function throughout life, the olfactory epithelium retains the capacity to repair itself to an extent not observable in other neuroepithelia or neuroepithelially derived tissue. The capacity for repair derives from the persistence of proliferating cells whose progeny become postmitotic and differentiate into neurons and other specialized elements of the epithelium. The stem cell elements reside within the poorly characterized population of globose basal cells, and their proliferation and differentiation are highly regulated normally. For example, damage to the olfactory nerve (e.g., due to head trauma) elicits retrograde neuronal degeneration and stimulates the proliferation of basal cells; in this setting, globose basal cells are fated to give rise to only neurons. After direct injury to the epithelium (e.g., by infection or by exposure to toxin), the stem cells become activated to replace both neurons and nonneuronal cells in the olfactory epithelium. Thus the epithelium has the capacity to reconstitute itself and restore function in this vital sensory system (**Fig. 40–3**).

Despite the capacity for reconstitution, when damage is severe enough to destroy stem cells that reside in the epithelium, the process of neuroepitheliopoiesis is aborted (**Fig. 40–3**). Thus, in such areas where the stem cells are lost, the epithelium reconstitutes as respiratory in appearance and function. In this instance the epithelial cells are largely derivatives of Bowman's gland elements. When significant percentages of olfactory epithelium are destroyed and replaced by respiratory epithelium, sensory loss ensues. The loss of sensory neurons is accompanied by the deposition of collagen within the fascicles of the olfactory nerve as a replacement for the preexisting axons.

APPROACH TO PATIENTS

Patient Presentation

Patients presenting with an olfactory complaint can be grouped by the degree of loss and the presence of altered sensations. Specifically, the complaints consist of

- Hyposmia—a partial loss of the ability to detect or identify odorants
- Anosmia—a total loss of the ability to detect or identify odorants
- Parosmia—the presence of sensation when an odorant is present, but the perceived quality is inappropriate for the stimulus
- Phantosmia—the presence of a sensation when no odorant is present
- Hyperosmia—a more acute than normal sense of smell

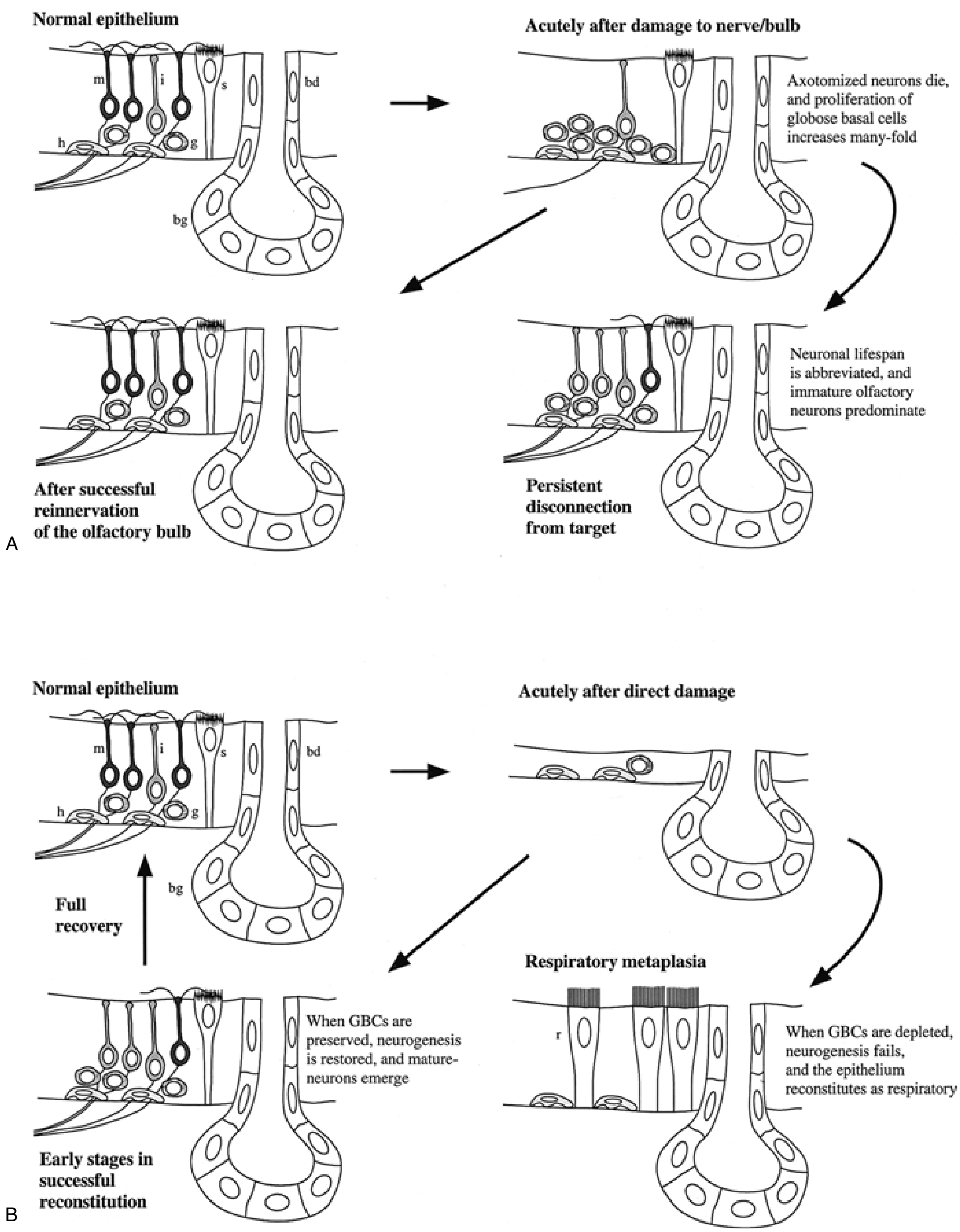


Figure 40–3 *Clinically relevant damage to the primary olfactory projection can take one of two forms, disruption of the olfactory nerve or destruction of the olfactory epithelium. Biopsies of the olfactory mucosa can be used to diagnose the form of damage and the extent of recovery.* ***(A)*** *In the case of nerve damage, successful restoration of the projection depends on the maintenance of an intact bulb and of the pathway to that target. With successful reinnervation of the bulb, the epithelium returns to its premorbid appearance. Persistent disconnection from the bulb results in accelerated turnover of olfactory neurons (i.e., an abbreviated lifespan). The reduction in lifespan has several effects. The rate of basal cell proliferation increases. The numbers of immature neurons increase in parallel; these can be identified by their expression of proteins related to axon growth and neuronal differentiation and the absence of cilia. Because lifespan is abbreviated, the newly born neurons have insufficient time to fully mature, and the population of mature neurons is markedly reduced; normally, these can be identified by their expression of the olfactory marker protein and the elaboration of cilia.* ***(B)*** *In cases where the olfactory epithelium is damaged directly; for example, by exposure to toxic substances, the outcome of the reconstitution process depends on the preservation of an adequate population of basal cells. If most or all of the basal cells are destroyed, the epithelium undergoes respiratory metaplasia, and it is repopulated by ciliated columnar respiratory epithelial cells, which transdifferentiate after migration away from the Bowman's glands and ducts. If an adequate population of basal cells is preserved, the epithelium will reconstitute as olfactory and recover to become indistinguishable from undamaged epithelium. Bowman's duct (bd); Bowman's gland (bg,); globose basal cell (GBCs); horizontal basal cell (h); immature neuron (i,); mature neuron (m,); ciliated respiratory epithelial cell (r); sustentacular cell (s).*

Hyposmia and anosmia represent different degrees of olfactory loss. When hyposmic, patients often complain that many odorants cannot be detected or that smells have to be stronger before they can be detected or identified. Because of the loss of smell, these patients often present with the complaint that food "doesn't taste as good" as it used to. In the literal sense, anosmic patients cannot detect and therefore cannot identify odorants. However, most odorants stimulate both cranial nerve (CN) I (perception of odor) and CN V (perception of irritation). Disruption of CN I does not preclude detection of odorant via CN V, although thresholds for odorant detection via CN V are generally much higher than for CN I. Because of the innervations of the nasal cavity by multiple branches of CN V, even direct facial trauma rarely eliminates all perception of nasal irritation. As a result, patients will often recognize and report that they can smell strong odors like ammonia or vinegar. Careful questioning and testing are necessary to distinguish between smelling the quality of odorants and detecting only the presence of irritation.

Parosmia and true parosmic shifts are characterized by a change in odorant quality. Patients may report that they can correctly identify the aroma of coffee but that the odor is not the same as before. Dramatic parosmic shifts are sometimes seen after olfactory loss due to head trauma, although many hyposmic patients also report minor parosmic shifts.

Phantosmia is the perception of an odor in the absence of an odorant. Phantom smells may be likened to tinnitus or pain from a phantom limb. They are most often putrid in quality, but a few patients report phantoms that are more pleasant. However, the presence of even a pleasant odor can become terribly distressing if it is omnipresent. Phantom smells can be constant or fluctuate in strength or duration. Sometimes they are dependent on nasal airflow and may be either uninasal or binasal. Phantom smells have been divided into peripheral (eliminated by application of a local anesthetic to the olfactory cleft, e.g., 1 to 2 mL of 4% topical cocaine) and central (not altered by peripheral anesthesia). Peripheral phantoms may be amenable to surgical treatment, whereas both central and peripheral phantoms may respond to therapy using antiepileptic drugs, including Neurontin (gabapentin), or antidepressants like Elavil (amitriptyline hydrochloride).

Hyperosmia is the report by a patient that he or she is more sensitive to odorants than in the past or in relation to other people. The reports can be quite reasonable, as in "I smelled the turkey cooking as soon as I turned onto my street," or unbelievable, as in "I could smell my neighbor lighting his cigarette even though both our windows were closed." The little work that has been done in this area has found patients suffering not so much from a shift in sensitivity as a shift in hedonic tone; that is, hyperosmic patients are much more bothered by odorants and tend to dislike them to a greater extent than do patients with a normal sense of smell. The cause for this hedonic shift is not currently known.

Anatomical Diagnosis: Where is the Lesion?

It is heuristically useful to consider olfactory dysfunction in anatomical terms analogous to diseases of the auditory system; that is, lesions can be classified as either conductive or sensorineural, depending on the disease process. The dysfunction can have a component of both (see below). Inflammation, infection, polyposis, and neoplasm of the lower nose often disrupt the transport of chemical stimuli by impinging on the narrow channel formed between the nasal septum and the conchae. Sensorineural lesions can be classified as peripheral versus central in nature. Examples include (1) sensory neuron destruction caused by direct damage to the olfactory epithelium (toxin or virus exposure), (2) damage to the olfactory nerve resulting in axonal transection and retrograde sensory neuron degeneration (head trauma), (3) neural lesions of the olfactory bulb and beyond (trauma and/or neurodegenerative disease), and (4) congenital abnormalities of the peripheral olfactory system due to maldevelopment of the CNS (examples include Kallmann's syndrome and the spectrum of holoprosencephaly-arrhinencephaly).

Finally, disruptions of the sense of smell can be a presenting complaint for intranasal neoplasms. These may arise from any part of the nose, including the olfactory epithelium. Olfactory neuroblastomas (also known as esthesioneuroblastomas) are locally aggressive small blue cell tumors that express proteins indicating a neural derivation. Thus the most likely cell of origin for such tumors are the globose basal cell precursors that proliferate at a brisk rate in normal epithelium.

Diagnostic Strategy

The evaluation of a patient with a chemosensory complaint includes the history, physical exam, chemosensory testing, and imaging studies. As with other areas of medicine, a thorough history is a valuable diagnostic aid; a few generalizations regarding the history follow. Given that the majority of olfactory disorders result from active local disease in the nose and paranasal sinuses, previous upper respiratory tract infections (postviral anosmia), or previous head trauma, it is important to inquire about these factors. Also, it is worth noting that chief complaints of a taste loss are often olfactory losses, which deserve substantiation by exam. A history of gradual loss and fluctuation of olfactory function is consistent with nasal/sinus disease. Other areas to explore include allergies, previous diagnostic procedures, surgeries, prior radiation therapy, systemic illness, and the review of systems. Present and prior medication use should be considered because many pharmaceuticals have been reported to be associated with chemosensory dysfunction (see **Table 40–1** for relatively well-documented examples).

The physical will include a full head and neck exam. Although direct anterior rhinoscopy may document mechanical obstruction or inflammation due to local disease processes, the olfactory mucosa lining the posterosuperior portion of the nasal cavity can only be visualized with endoscopy. Inspection using rigid or flexible fiberoptic endoscopy should assess patency, polyps, tumors, septal deviation, and mucosal thickening, as well as purulence and mucus quality. Imaging studies are necessary to rule out intracranial pathology and to further assess the anatomy of the peripheral olfactory system. Coronal computed tomographic (CT) scans are ideal for examining the nasal air space and paranasal sinuses and may reveal mucosal thickening, the presence of fluid levels, or masses. Additionally, bone algorithms will allow assessment of the cribriform plate region, through which the olfactory nerves must traverse to synapse with the olfactory bulbs. Because of the prevalence of nasal and sinus disease, evidence of its presence in a patient with an olfactory complaint does not necessarily imply a causal relationship; other pathology should still be considered. The otolaryngologist may entertain the notion of taking biopsies in the olfactory area, to evaluate the status of the olfactory epithelium, and of the fascicles of the olfactory nerve. In particular, the extent to which axons have been replaced by collagen in the nerve provides an integrated assessment of the extent to which olfactory epithelium has been replaced by respiratory epithelium. Finally, at least rudimentary assessment of olfactory function must be performed if more detailed testing is not feasible.

Psychophysical Evaluation

Great progress has been made over the past 15 years in standardizing clinical olfactory testing. These include tests of odorant threshold, which can be likened to standard audiometric evaluation of auditory threshold, and tests of odorant identification, which can be likened to evaluation of speech recognition.

Odorant Threshold

Operationally, *threshold* is defined as the lowest concentration of a particular odorant that can be detected reliably. Generally, testing consists of a two-interval forced-choice procedure embedded in either the method of limits or a tracking procedure. In both techniques, the patient is presented with two bottles, one containing the odorant and the other a nonodorous solvent. The task of the patient is to pick which of the two bottles contains the odorant. If the patient answers incorrectly, the concentration of the odorant is elevated either two- or threefold, and the procedure is repeated. If the patient answers correctly, the odorant concentration remains the same, and the procedure is repeated. In the method of limits, the odorant concentration at which five consecutive correct answers are given is defined as threshold. In the tracking procedure, two consecutive correct answers result in the presentation of a subsequent odorant pair at the next lower concentration. As such, the odorant concentrations presented will vary around threshold. Threshold is defined as the average of the last four reversal points in a threshold series in which the odorant presentation has switched between increasing and decreasing seven times. In both procedures, care should be taken to include different blanks with each odorant concentration to reduce extraneous cues in the detection of the odorant. In addition, presentation of the blank and the odorant should be randomized or counterbalanced so as not to aid the detection process.

Care must be taken in the interpretation of threshold elevation. Like audition, where a high-end hearing loss does not necessarily imply a low-end loss, elevation of threshold for one odorant does not necessarily imply elevation for all odorants. Having said this, an elevation of the threshold of one odorant often implies elevation for other odorants.

Odorant Identification

In tests of odorant identification, the patient is presented with a series of odorants and asked to identify each odorant, usually from a list of odorant names. By far the most popular of these tests is the Smell Identification Test, or SIT (Sensonics Inc., Haddon Heights, NJ). This test consists of 40 scratch-and-sniff odorant identifications. Each microencapsulated odorant is accompanied by four possible odorant names, one of which is correct. The patient must choose which one of the four odorant names best represents the odorant. The test includes extensive normative data for men and women ages 5 to 90. The test–retest reliability is excellent, and the test has

TABLE 40–1 CONDITIONS ASSOCIATED WITH OLFACTORY DYSFUNCTION AND THE LOCUS OF THE LESION

	Sensorineural				
Conductive	**Epithelium-Based**	**Nerve-Based**	**Bulb-Based**	**Beyond the Bulb**	**Locus-Uncertain**
Structural abnormalities of the nose	Specific hyposmias		Congenital/hereditary: Kallmann's syndrome Holoprosencephaly spectrum	Epilepsy	Nutritional/metabolic (possibly OE-based)
Nasal trauma		Head trauma (coup and contracoup)	Head trauma Surgical ablation	Head trauma	Drugs (possibly OE-based)
Infectious rhinosinusitis: Viral Bacterial Fungal	Infection-mediated destruction of OE: Viral URI Bacterial rhinosinusitis		Infection: viral Encephalitis via ON HSE Rabies (bat guano) Picornavirus (?)		
Inflammatory rhinitis: Granulomatosis Autoimmune disease Vasomotor Allergic Rebound Polyposis	Toxins and environmental agents: Industrial dusts Inorganic pollutants Organic solvents X-irradiation	Disordered axonal growth (idiopathic)	Disordered re innervation after massive damage to OE		
Neoplasms	Neoplasms: Olfactory neuroblastoma Aging Neurodegenerative diseases (?)	Neoplasms: Schwannoma Neurofibroma	Neoplasms: Olfactory groove meningioma Neurodegenerative diseases: Alzheimer's Parkinson's	Neoplasms: Glioma of first- or second-order olfactory cortical areas in frontal and temporal lobe Neurodegenerative diseases: Alzheimer's Parkinson's	

HSE, *Herpes simplex* encephalitis; OE, olfactory epithelium; ON, olfactory nerve; URI, upper respiratory tract infection.

been shown to be sensitive to the olfactory loss that accompanies many different disease processes.

Many other tests of olfactory function are possible, including odorant recognition threshold, odorant intensity perception, and odorant memory, to name a few. However, currently, these tests are not in wide use clinically.

It should be cautioned that, although tests of odorant threshold and identification have proven useful in the quantification of olfactory loss associated with many disease processes, these tests have not provided assistance in distinguishing between losses that are conductive (inability of odorant molecules from reaching the olfactory receptors) and those that are sensorineural (neuronal damage). For example, a loss of odorant identification could be the result of nasal congestion, receptor damage, damage to central structures, or even disruption of working memory or memory stores. One test that holds some promise for making this distinction is the Odorant Confusion Matrix, or OCM (developed at SUNY–Syracuse, NY). Like the SIT, the OCM is an odorant identification task, although it is structured very differently. The patient is asked to identify 10 different odorants, and although percent of correct identification gives a measure of olfactory loss, the errors or confusions in odorant identification may provide a clue to the underlying mechanism of loss. Certainly, more work is necessary in this area before tests of olfaction can provide assistance in the sorting out of possible etiologies of olfactory loss.

The vast majority of odorants stimulate both the olfactory and the trigeminal nerves. Although trigeminal thresholds are generally higher and often much higher than olfactory thresholds, a good part of normal odorant perception results from the combination of these two neural inputs. Because of this, many patients report the ability to perceive "strong" smells. It is important to distinguish between the ability to perceive the quality associated with these odorants and the irritation. Most anosmic subjects can easily detect the presence of ammonia and vinegar but cannot distinguish between these qualities except by inferences from the intensity of the irritation. Careful histories can help to sort out these questions and allow an estimation of whether the patient is anosmic with some trigeminal function versus hyposmic.

Often clinicians are asked to make legal judgments as to the veracity of claims of olfactory loss. It is especially important in cases of litigation that the results of olfactory testing be accurate to facilitate possible appropriate medical treatment. The detection of malingering basically takes two forms. The first is inference from statistical probability theory. As an example, people asked to feign an olfactory loss often perform more poorly on a given test than would be suggested by chance alone. The only way to perform this poorly is if the correct answer is known, an outcome that is only consistent with an excellent olfactory ability. The second method is to use the common lack of understanding concerning the dual sensitivity of the olfactory and trigeminal nerves to most odorants. True anosmics will readily report that they can detect strong odorant such as ammonia. However, people who are asked to malinger usually treat all odorants the same and fail to report the presence of either odorants or irritants. This effect is especially apparent in the results from the OCM. Because of the design of the test, anosmic patients will often confuse ammonia with vinegar (both odorants with a strong trigeminal component) but will rarely confuse the smell of these chemicals with names of odors that are not associated with perceptual irritation. For example, an anosmic subject would rarely confuse vanilla for ammonia because, by personal experience, vanilla is not associated with a sting. Patients feigning an olfactory loss are not aware of this fact and will use both irritant and odorant names for chemicals with obvious nasal sting.

ETIOLOGIES/PATHOPHYSIOLOGICAL MECHANISMS

Within the two general categories of olfactory dysfunction—sensorineural loss versus conductive block—exist different proximate causes and different mechanisms of disease. **Table 40–1** outlines the various types of disease processes and the location the process affects that can give rise to olfactory dysfunction, whether anosmia, hyposmia, parosmia, or phantosmia. There are several important points to make with reference to the specific items in the table. First, some disease processes can cause olfactory dysfunction by affecting both the conduction of stimuli to the olfactory area and transduction by the sensory apparatus. For example, an upper respiratory tract infection (URI) can produce mucosal inflammation, hampering delivery of the chemical stimuli to the olfactory area and their transfer into the mucus in which the sensory apparatus is embedded. In addition, viral infection and inflammation can acutely cause destruction of olfactory epithelium, including sustentacular cells, neurons, and basal cell progenitors. Second, damage to the olfactory epithelium, if sufficiently severe, will cause disease despite the capacity for regeneration of the neuronal population, as already noted (**Fig. 40–3**). The failure of neural regeneration derives from the destruction of the population of globose basal cells in the affected area and results in respiratory metaplasia. If the replacement of olfactory epithelium by respiratory epithelium is sufficiently extensive, hyposmia or even anosmia can result. Third, even after successful reconstitution of the olfactory epithelium, reinnervation of the olfactory bulb can go awry. As noted earlier, sensory function depends on the highly ordered connectivity between olfactory epithelium and olfactory bulb; two chemical stimuli are distinguishable when they elicit different patterns of activity across the glomerular surface of the olfactory bulb. This glomerular activation pattern is a direct result of the activation of particular members of the odorant receptor family (i.e., a particular set of olfactory sensory neurons). The differential patterns of bulbar activity elicited in response to behaviorally distinguishable stimuli emerge from the stimulation of different subsets of the family of odorant receptors. Total or near-total destruction of the olfactory neurons eliminates preexisting axons, which help guide axons of like sensory neurons to the appropriate glomeruli under normal conditions. Even when reconstitution returns the olfactory epithelium to normal or near normal, near-complete lesions have the consequence that the specificity with which olfactory axons from newly born neurons reinnervate the olfactory bulb is degraded. In other words, nascent axons are mistargeted after massive destruction in the periphery, which has the effect of altering the conversion of chemical structure to neural space code. Thus a not uncommon complaint of patients is observed: "I know what's there [gasoline, perfume, citrus fruit], but it doesn't smell like it used to." Fourth, in addition to the misdirection of axons that reach the olfactory bulb, disordered growth can occur within olfactory fascicles. As a consequence, neuromas form in both olfactory epithelium and within the fascicles per se. Neuroma formation is observed in all patients with the complaint of phantosmia. Ephaptic transmission occurs at sites of neuroma formation in the peripheral nervous system and underlies symptoms of causalgia after peripheral nerve damage. A similar mechanism may operate in phantosmic patients to compromise the fidelity of sensory transmission.

THERAPEUTIC STRATEGIES

Conductive Disease Processes

For conductive problems, in which mechanical obstruction impedes airflow to the olfactory mucosa, the goal of therapy is the relief of obstruction and elimination of persistent infection. Common examples of this process are nasal polyposis and chronic sinusitis with hyposmia or anosmia. In terms of medical therapy, it is reasonable to treat any infection initially and to attempt to relieve obstruction. Antibiotics followed by topical corticosteroids (e.g., flunisolide) for a minimum of 8 weeks have been reported to be effective. Should medical therapy fail, surgical treatment with functional endoscopic sinus surgery may be considered (see Chapter 38 for a discussion of surgical anatomy). In addition, patients with adenoidal hypertrophy or septal deviation may derive benefit from appropriate surgical treatment addressing these problems.

Sensorineural Disease Processes

At present, the therapeutic armamentarium that is available for sensory neuronal problems is extremely limited. Beyond identifying treatable factors—side effects of prescription medicines, ongoing exposure to toxins, central or peripheral neoplasms, and chronic infections—little can be done presently, other than counseling for the difficulties of daily life without a sense of smell. The counseling aspect should not be neglected, however. Loss of smell can affect the types of food patients may choose, and their appetite may either increase or decrease, necessitating some nutritional counseling. Smell impairment also will require special attention to substituting other means of warning patients to the presence of dangerous situations, such as natural gas, smoke/fire, and spoiled food, that are normally conveyed by olfactory cues. Accommodations to these concerns include the installation of smoke alarms and gas detectors and the careful labeling and dating of food. A smell loss may have other effects on a patient that are more closely aligned with quality of life (food, friends, personal hygiene, and the smell of a spring day). Although some patients manage despite these daily problems, others find themselves in need of additional professional help.

One exciting exception to this therapeutic nihilism relates to patients complaining of symptoms of phantosmia. In some of these patients, symptomatology can be reduced if not eliminated by treatment with antiepileptic or antidepressant medication. In other patients whose phantosmia is debilitating but refractory to drug therapy, the symptomatology can be alleviated by surgical extirpation of the olfactory mucosa via an endoscopic approach. Alternatively, surgical ablation of the olfactory bulb after frontal craniotomy, which is a significantly riskier solution, has been used to alleviate the symptomatology. Certain clinical criteria must be satisfied before taking the drastic step of ablating the mucosa. The phantosmic symptoms must be chronic (greater than 6 months' duration), elicitable by the patient or physician, profoundly distressing, and persistent despite an adequate trial with antiepileptic and/or antidepressant medication.

The patient must be evaluated psychiatrically, and the usual surgical risks need to be eliminated. In addition, it is essential for the physician to demonstrate that anesthetization of the olfactory mucosa or blocking access of stimuli to the olfactory cleft alleviates the symptomatology.

Future Treatment Strategies

The field of rhinology continues to grow. With the expansion of the geriatric population, the numbers of patients with olfactory dysfunction will continue to increase. As we learn more about the basic biology of the olfactory system, novel treatment strategies may emerge. The phenomenon of patchy replacement of the olfactory neuroepithelium with respiratory epithelium is well documented, increases with age, and is likely to contribute to olfactory dysfunction. Thus, future therapeutic modalities will seek to foster regeneration of olfactory epithelium after severe peripheral lesion, either by replacing stem cells via transplantation or stimulating stem cell proliferation through the action of the appropriate growth factors and cytokines. Unfortunately, our understanding of the networks and cascades that regulate stem cell behavior is far too primitive at the present time to propose specific therapeutic regimens of either nature. Additionally, strategies to address the problem of scarring or axon obstruction at the cribriform plate, important in the process of post-traumatic anosmia, are of interest. Finally, better understanding of the allergy and immunology of the respiratory mucosa will lead to new strategies for the common causes of mechanical obstruction with dysosmia.

ACKNOWLEDGMENTS

Preparation of the chapter was supported by grants from the NIH P01 DC02220, R01 DC00467, and R01 DC02167.

Suggested Readings

Doty RL, ed. Handbook of Olfaction and Gustation. New York: Marcel Dekker; 1995

Farbman AI. Cell Biology of Olfaction. Cambridge: Cambridge University Press, 1992

Getchell TV, Doty RL, Bartoshuk LM, Snow JB Jr, eds. Smell and Taste in Health and Disease. New York: Raven Press; 1991

Leopold DA, Schwob JE, Youngentob SL, Hornung DE, Wright HN, Mozell MM. Successful treatment of phantosmia with preservation of olfaction. Arch Otolaryngol Head Neck Surg 1991;117:1402–1406

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. A viral upper respiratory tract infection can cause olfactory dysfunction by
   - A. Blocking conduction of stimuli to the olfactory cleft due to mucosal inflammation
   - B. Acutely destroying olfactory epithelium
   - C. Causing replacement of olfactory epithelium by respiratory epithelium, when severe
   - D. All of the above
2. Many chemicals, for example, pungent stimuli like vinegar, activate both trigeminal nerve fibers and olfactory sensory neurons
   - A. True
   - B. False
3. A patient whose complaint of being unable to smell is due to dysfunction in cranial nerve I (the peripheral olfactory system) will be able to detect (i.e., have a measurable threshold for)
   - A. Ammonia
   - B. Vanilla
   - C. Neither of the above
4. A patient's complaint of frequent but intermittent phantosmia is blocked by anesthetization of the olfactory epithelium, indicating that
   - A. The aberrant perception requires stimulation of the olfactory epithelium
   - B. The phantom percept is exclusively mediated by disease of the peripheral sensory apparatus
   - C. The patient potentially would benefit from an ablative procedure
   - D. A and C
   - E. All of the above

Part IV

THE LARYNX, VOICE, AND NECK

Chapter 41

THE BRANCHIAL ARCHES AND THEIR DERIVATIVES

Jeffrey T. Laitman, Joy S. Reidenberg, Armand Balboni,
Andrew Bergemann, and Peter Som

The embryonic branchial arches and their derivatives are major structures in establishing the anatomy of the head and neck. An understanding of the developmental relationships among the derivative structures is essential to proper clinical treatment and subsequent care. Indeed, failure to understand the developmental underpinnings among arch derivatives would deprive a clinician of needed diagnostic information.

The branchial arches are the main component of the branchial "apparatus," which includes the branchial arches, pharyngeal pouches, branchial clefts, and branchial membranes. The major components in this group are the branchial arches, transient embryonic structures from which will arise many of the postnatal structures of the head and neck. The word *branchial* derives from the Greek *branchia*, meaning "gills." Early in mammalian development, the cranial region of an embryo has marked archlike regions with intervening clefts. To early anatomists, this condition resembled the gills of a fish (they are, in fact, homologous to the precursors of gills in fish and immature amphibians), and hence the term *branchial arch* came into common use. Humans never develop gills, and as such the term *pharyngeal arch* is frequently used. We will use the term *branchial arch* throughout this chapter because it is the more commonly used term in most basic and clinical studies, and it is also used when describing these structures in most studies of vertebrates other than humans.

Different terms are often applied to the development of head and neck structures and the time frames within which this development occurs. As such, it is important to establish briefly those terms that we will use in this chapter. First, the terms *embryo* and *embryogenesis* are properly used to refer to the first 8 weeks of development in humans. This period largely corresponds to the appearance of organs and organ systems. The term *organogenesis* thus is largely equated with the "embryonic" period. Following the embryonic period is the "fetal" period. This is largely a time of growth and differentiation as opposed to the appearance of structures. Accordingly, one should not refer to an individual before the end of the eighth week as a fetus, nor after the ninth as an embryo. It should be noted that these demarcations, terms, and time frames are based on humans. Gestational periods differ for other animals, and so the periods and terminology may not mirror those of humans. This is important to keep in mind in studies of comparative embryology in general, in which several rodent or avian species are used as experimental models.

EMBRYOLOGY OF THE BRANCHIAL APPARATUS

The branchial apparatus is first identified about the fourth week of development and is completed by the sixth to seventh week. Initially, ventrally migrating neural crest cells interact with lateral extensions of the ventral pharyngeal endoderm that surrounds the six paired aortic arch arteries. This appears to initiate branchial arch development and results in the segmentation of the mesoderm lateral to the ventral foregut to form a series of five distinct bilateral pairs of mesenchymal swellings, referred to as the branchial arches. The initial mesodermal core of each arch is then augmented by migrating neural crest cells that surround the mesodermal elements. This mesoderm will give rise to muscle myoblasts, and the neural crest cells will give rise to skeletal and connective tissues.

Externally, the ectodermally covered branchial arches are separated by branchial clefts or grooves, which open into invaginated pouches. Each branchial cleft/groove is associated with a corresponding internal pouch derived from the endodermally lined foregut, referred to as a pharyngeal pouch. The external clefts and internal pouches abut each other and correspond to the separations between the arches. Only four paired branchial clefts appear externally, but five paired pharyngeal pouches develop internally. The fifth arch does not appear on the surface but instead lies buried about the site of origin of the laryngotracheal outgrowth. Likewise, the fifth pouch is only represented by a transient invagination.

The existence of the elusive “fifth” arch is itself an often debated topic because it is unclear as to whether the fifth arch ever develops in humans. Because of this confusion, the fifth arch is often placed together with the sixth arch, and its derivatives are thus perhaps more conservatively described as originating from the “fifth-sixth” arches. This terminology will be used here.

Thus there are five pairs of identifiable arches that are numbered cranially to caudally as the first, second, third, fourth, and fifth-sixth branchial arches. The branchial arches decrease in size from cranial to caudal, and each pair is united midventrally. These arches are prominent in a lateral profile of the embryo, and they are aligned transversely to the long axis of the neck.

The cleft between each branchial arch forms a thin double-layered branchial or closing membrane where it transiently comes in contact with the endoderm of the primitive pharynx. However, mesoderm soon separates the ectodermal and endodermal layers of the membranes. The tympanic membrane (eardrum), which is derived from the first branchial membrane, is the only branchial membrane structure to remain as such in the adult human. It should be noted that the clefts never communicate with the foregut lumen as they do in the gill apparatus of the fish. The cleft and pouch (and thus the branchial membrane) of each arch are positioned caudally to the arch (mesoderm) of the same number. Each arch contains a central cartilaginous rod that is differentiated from neural crest tissue and that forms the skeleton of the arch. This tissue is destined to become bone, cartilage, or ligamentous structures. There is also a muscular component, an artery that runs around the developing pharynx from the ventral heart to the dorsal aorta, and a neural element that supplies the mucosa and muscles that will arise from that arch. These neural elements consist of sensory and specialized visceral motor fibers of one or more cranial nerves. It is the migrating cranial nerve fibers of each arch that initiate the development of that arch’s muscles; although some of these muscles may migrate from their sites of origin, these muscles retain their original arch nerve supply. It is also the migration of the muscles, which “drag” their neural supply with them, that accounts for the tortuous routes of many of the cranial nerves.

Shortly after the branchial arches appear, excessive mesodermal growth occurs in the first arch, the cranial portion of the second arch, and the epipericardial ridge, which develops from the mesoderm lateral to the fifth-sixth arch. As the second arch expands laterally, it also curves caudally, overlapping the external surfaces of the third and fourth arches until it reaches and fuses with the external surface of the sixth arch and cardiac eminence. In so doing, it obliterates the outer contour of the second through the fourth grooves and forms the hyoid operculum. These accelerated growths result in a submerging of the intervening caudal portion of the second, third, and fourth arches and associated branchial clefts into a shallow, recessed ectodermal pit, the retrohyoid depression, or the lateral cervical sinus of His. Further growth about the cervical sinus results in narrowing of the external opening into a channel called the cervical duct. Soon the cervical duct is obliterated, as is the ectodermally lined cervical sinus, and eventually there is a smooth, uniform contour to the external surface of the neck. In the adult, the site of the cervical sinus is located at the angle between the dorsal surface of the supra- and infrahyoid "strap" muscles and the anterior margin of the sternocleidomastoid muscle.

BRANCHIAL ARCH DERIVATIVES

As mentioned earlier, specific osseous, cartilaginous, ligamentous, nervous, and vascular structures arise from each branchial arch (**Table 41–1**). However, the muscles of the different arches do not necessarily attach to the osseous or cartilaginous components of their own arches. In some cases, the muscles may migrate to the surrounding regions. The origins of these muscles can always be established through their nerve supply, which comes from their arch of origin.

The first, or mandibular, pair of branchial arches gives rise to the precursors of the jaws. On each side of the face, the first arch borders the lateral margin of the stomodeum. The maxilla is derived from the small maxillary prominence that extends ventrally from the much larger mandibular prominence. The cartilage skeleton of the first arch is Meckel's cartilage, which can be identified between 41 and 45 days of gestation. Most of this cartilage disappears in the formed mandible. The smaller dorsal portion of this cartilage is related to the developing ear and eventually becomes ossified to form the upper portions of the malleus and incus. The intermediate portion of this cartilage regresses, and its perichondrium forms the anterior ligament of the malleus and the sphenomandibular ligament. The ventral portion of this cartilage largely disappears, and the mandible is formed around it by secondary intramembranous ossification.

The trigeminal or fifth cranial nerve (CN V) is the nerve of the first arch, and it innervates the muscles of the first arch, which include the muscles of mastication (medial and lateral pterygoids, masseter, and temporalis), mylohyoid, anterior belly of the digastric, tensor tympani, and tensor veli palatini muscles. In addition to this motor supply, the trigeminal nerve provides sensory supply to the region of the first branchial arch. This includes the mandible, its covering mucosa and gingiva, the mandibular teeth, the mucosa of the anterior two thirds of the tongue, the floor of the mouth, and the skin of the face. The first arch artery contributes to a terminal branch of the maxillary artery.

Reichert's cartilage, the cartilage of the second branchial or hyoid arch, appears between 45 and 48 days of gestation. The dorsal margin of the second arch is also closely related to the middle ear and ossifies to form the manubrium of the malleus; the long process of the incus; the head, neck, and crura of the stapes; and the styloid process of the temporal bone. The footplate (base) of the stapes is not a branchial arch derivative because it is derived from the otic capsule. The portion of the cartilage between the styloid process and the hyoid bone regresses,

TABLE 41–1 KEY DERIVATIVES OF BRANCHIAL ARCHES

Branchial Arch	Muscles	Cranial Cerve (CN)	Aortic Arch Artery	Skeletal Elements
1	Muscles of mastication, mylohyoid, anterior belly digastric, tensor tympani, tensor veli palatini	Trigeminal (CN V)	Maxillary	Meckel's cartilage (around which mandible will form): upper portions of malleus and incus, sphenomandibular ligament
2	Muscles of facial expression, posterior belly digastric, stylohyoid, stapedius	Facial (CN VII)	Stapedial (not present postnatally)	Reichert's cartilage: stapes (except footplate), styloid process, stylohyoid ligament, lesser horns and upper body of hyoid
3	Stylopharyngeus	Glossopharyngeal (CN IX)	Common carotid, root of internal carotid	Greater horns and lower body of hyoid
4	Cricothyroid, palatofaucial, and rostral pharyngeal	Vagus (CN X)	Aortic arch, right subclavian, brachiocephalic	Laryngeal cartilages
5/6	Intrinsic muscles of larynx (except cricothyroid) Caudal pharangeal	Cranial root of spinal accessory (CN XI)	Ductus arteriosus, roots of pulmonary arteries	Laryngeal cartilages

and its perichondrium forms the stylohyoid ligament. (It is interesting to note that many adult mammals exhibit a chain of bony elements spanning between the styloid process and hyoid bone.)

The ventral end of Reichert's cartilage ossifies to form the lesser cornua (horns) and the upper portion of the body of the hyoid bone. The muscles of the second arch include the posterior belly of the digastric, the stylohyoid, the stapedius, and muscles of facial expression, all of which are supplied by the nerve of the second branchial arch, the facial nerve (CN VII). Other than a small sensory branch of CN VII, which may supply a portion of the external auditory canal, there is no sensory distribution from CN VII to the ectodermal derivatives. The main sensory component of the facial nerve, the chorda tympani, invades the first arch as a pretrematic nerve (i.e., a nerve that joins the nerve of the arch above by traveling over the trema, or interarch, cleft). The chorda tympani is thus carried in a branch of the trigeminal nerve to supply taste to the mucosa of the anterior two thirds of the tongue. The artery of the second arch is the stapedial artery, which disappears during the fetal period. Its prior location is marked by the stapes' central foramen, the gap between the stapes crura. Portions of the stapedial artery may persist as the corticotympanic artery in the adult.

The cartilage of the third branchial arch ossifies to form the greater cornua and lower portion of the body of the hyoid bone. The musculature of this arch is limited to the stylopharyngeus muscle, which is innervated by the nerve of the third arch, the glossopharyngeal nerve (CN IX). Because the mucosa of the posterior one third of the tongue is also derived from the third branchial arch, its sensory innervation is supplied largely by the glossopharyngeal nerve. Neural crest tissue in the third arch forms the carotid bodies, which appear as mesenchymal condensations about the third aortic arch artery. Thus the glossopharyngeal nerve (CN IX) supplies innervation to the carotid body. The artery of this arch contributes to the common carotid artery and part of the root of the internal carotid artery.

The cartilages of the fourth and fifth-sixth branchial arches fuse to form the larynx, although the precise arch of origin for some of the laryngeal structures clearly remains uncertain. Overall, it is believed that the line of division between the fourth and sixth arches is the vocal folds (true vocal cords). It is generally thought that all laryngeal cartilages (with the possible exception of the epiglottis) are derived from the fourth through sixth arches. It would appear logical that the cartilages are laid out in a craniocaudal arrangement corresponding to their arches of origin, but this remains to be shown. The cricothyroid muscle of the larynx and portions of the pharyngeal muscles (most probably the more rostral components) are derived from the fourth branchial arch. They are innervated by the nerve of this arch, the vagus nerve (CN X, particularly the superior laryngeal branch), which also gives sensory innervation to the mucosa of the larynx above the vocal folds. The muscles of the fifth-sixth branchial arches are the remaining intrinsic muscles of the larynx (exclusive of the cricothyroid) and portions of the pharyngeal muscles (the more caudal muscles, particularly the inferior constrictor) innervated by the nerve of the fifth-sixth arch. This nerve is generally believed to be the cranial portion (i.e., "cranial root") of the spinal accessory nerve (CN XI), whose fibers join the vagus and are distributed within its recurrent laryngeal branch. The striated muscles of the upper half of the cervical esophagus are also derived from the sixth arch. The caudal portion of the esophagus is composed of smooth muscle and is derived from the splanchnic mesoderm of the primitive foregut. In addition, the para-aortic bodies of chromaffin cells that secrete noradrenaline arise from the ectomesenchyme of the fourth and sixth arches.

The left artery of the fourth arch forms the aortic arch. The right arch artery contributes to the brachiocephalic and right subclavian arteries. They also form the original sprouts of the pulmonary arteries. The sixth arch arteries develop in part into the roots of the definitive pulmonary arteries. The remainder of the artery disappears on the right side; on the left side it forms the fetal ductus arteriosus that in the adult becomes the ligamentum arteriosum.

BRANCHIAL CLEFT AND PHARYNGEAL POUCH DERIVATIVES

Of all the paired branchial clefts (branchial groove, pharyngeal cleft), only the first cleft contributes directly to a postnatal structure. It persists as the external acoustic meatus and the epithelium of the external auditory canal. The first pharyngeal pouch forms the tubotympanic recess in the fourth week, which becomes the tympanic cavity and auditory (pharyngotympanic or eustachian) tube in the fifth week, thereby connecting the left and right middle ear spaces with the pharynx. The membrane remaining between the first branchial cleft and the first branchial pouch becomes the tympanic membrane (eardrum) by the end of the fifth week. The other branchial clefts, together with the cervical sinus of His, normally are obliterated as the neck develops.

Unlike the branchial clefts, all the remaining paired pharyngeal pouches give rise to important postnatal structures. Pouch 2 gives rise to the stroma of the paired palatine tonsils in the second trimester, forming hollow tonsillar crypts that are then infiltrated by lymphoid tissue to form lymph follicles in the third trimester. Pouch 3 forms the thymus beginning in the fourth week. Thymic tissues from the left and right sides migrate caudally and fuse to each other in a midline position inferior to the thyroid gland to form the single thymus gland, which is later infiltrated by lymphocytes by the end of the first trimester. In the fifth week, the inferior parathyroid gland arises from pouch 3 (and is thus sometimes called parathyroid III), and the superior parathyroid gland arises from pouch 4 (and is thus sometimes called parathyroid IV). Although both sets of paired parathyroid glands migrate caudally into the thyroid gland by week 7, the more proximally arising inferior parathyroids migrate further caudally than the more distally arising superior parathyroids. Thus their names (superior vs inferior) reflect only the final relative positions of these glands within the thyroid gland. The origin of the paired ultimobranchial (telopharyngeal) bodies is controversial. Their origin appears to be related to either the caudal portion of pouch 4 or a transient invagination believed to be a remnant of pouch 5 that appears in the fifth week. This invagination is populated by epithelial cells that migrate caudally and implant into the thyroid gland, where they then differentiate into parafollicular cells (calcitonin-producing C cells).

TONGUE

The covering of the tongue develops from the first, third, and fourth branchial arches, and the intrinsic musculature develops from the occipital somites. The first arch gives rise initially to the median tongue bud (tuberculum impar) in the fourth week, which is later overgrown by lateral lingual swellings (right and left distal tongue buds) arising from this same arch beginning in the fifth week. These first arch swellings form the anterior two thirds of the tongue's surface. This region is innervated for sensation by cranial nerve branches of the first and second branchial arches: the lingual branch of the mandibular division of the trigeminal nerve (V_3) for touch, and the chorda tympani from the facial nerve for taste carried in V_3.

The copula is a transient midline swelling of the second arch in the fourth week that is later overgrown by the hypobranchial (hypopharyngeal) eminence, a midline swelling of the third and fourth arches that develops in the fifth and sixth weeks. The hypobranchial eminence forms the posterior third of the tongue's surface and is innervated for sensation by the glossopharyngeal nerve for touch and taste, and by the vagus for taste at its most posterior edge near the epiglottis. Myoblasts from occipital somites form the intrinsic musculature of the tongue and are innervated by a nonbranchiomeric cranial nerve, the hypoglossal (CN XII).

The boundary between the first and third arch contributions is marked by a transverse groove called the terminal sulcus, and the midline groove separating the right and left distal tongue buds is called the median sulcus. The point of intersection between the terminal and median sulci marks the location of the foramen cecum. This is the site where the primordium of the thyroid gland appears in the second through fourth weeks. The descending thyroid gland leaves behind a thyroglossal duct, which normally obliterates by the end of the fifth week. The thyroid gland then descends caudally to its final position immediately ventral and inferior to the larynx.

SUMMARY

The branchial apparatus contributes significantly to the structures of the head and neck. This apparatus is composed of four components: the branchial arches, pharyngeal pouches, branchial clefts, and branchial membranes. The major component of this group is the branchial arches, transient structures from which derive major muscular, nervous, vascular, and skeletal elements. Some of the major features in the development of these structures are highlighted in this section.

The first branchial arch mesenchyme gives rise to the muscles of mastication and some of the swallowing muscles. Swellings of the first arch contribute to the covering of the anterior two thirds of the tongue. The cranial nerve of the first arch is the trigeminal, and the associated artery is the maxillary. The skeletal elements of the first arch derive from the transient Meckel's cartilages and include the sphenomandibular ligament and the upper portions of the malleus and incus. The first branchial cleft forms the external acoustic meatus and ear canal, and the first pharyngeal pouch forms the tympanic cavity and pharyngotympanic (auditory) tube. The intervening branchial membrane forms the tympanic membrane.

The second arch mesenchyme gives rise to the muscles of facial expression. The cranial nerve of the second arch is the facial, and the associated artery is the transient stapedius. The skeletal elements of the second arch derive from Reichert's cartilage and include the lower portions of the malleus and incus, stapes (not the footplate), styloid process of the temporal bone, stylohyoid ligament, and upper portion of the body and the lesser horns of the hyoid. The second pharyngeal pouch forms the crypts of the palatine tonsils.

The mesenchyme of the third, fourth, and fifth-sixth arches merge into the palatofacial, pharyngeal, and laryngeal muscles, and contribute to the covering of the posterior one third of the tongue, the pharynx, and the larynx. The cranial nerve of the third arch is the glossopharyngeal, and the associated arteries are the common carotid and root of the internal carotid. The skeletal elements of the third arch are the lower portion of the hyoid body and the greater cornua. The third pouch gives rise to the thymus and the inferior parathyroids.

The cranial nerve of the fourth arch is the vagus (particularly the superior laryngeal branch), and the cranial nerve of the fifth-sixth arches is the cranial root of the spinal accessory nerve, which is carried in the recurrent laryngeal branch of the vagus. The fourth pair of arch arteries forms the adult aortic arch on the left side and the brachiocephalic and right subclavian on the right side. The sixth arch arteries develop into the roots of the pulmonary arteries and on the left side also form the fetal ductus arteriosus (adult ligamentum arteriosus). The fourth and fifth-sixth arches contribute to the cartilages of the larynx. The fourth pouch gives rise to the superior parathyroids and (possibly in combination with a fifth pouch) the ultimobranchial body.

Suggested Readings

Cobourne MT. Construction for the modern head: current concepts in craniofacial development. J Orthod 2000;27(4):307–314

Graham A. The development and evolution of the pharyngeal arches. J Anat 2001;199(pt. 1–2):133–141

Graham A, Smith A. Patterning the pharyngeal arches. Bioessays 2001;23(1):54–61

Mandell DL. Head and neck anomalies related to the branchial apparatus. Otolaryngol Clin North Am 2000;33(6):1309–1332

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. The existence of which of the branchial arches in humans remains unclear and thus controversial?
   A. First and second
   B. Third
   C. Fourth
   D. Fifth
   E. Sixth
2. Skeletal or cartilaginous derivatives of the second branchial arch include
   A. Meckel's cartilage
   B. The head of the malleus
   C. The head, neck, and crura of the stapes but not the footplate
   D. The greater horns and lower portion of the body of the hyoid bone
   E. The thyroid cartilage
3. What is a postnatal derivative of the first branchial cleft?
   A. The footplate of the stapes
   B. The external acoustic meatus
   C. The cervical sinus (of His)
   D. The incudomalleolar joint
   E. The middle ear cavity
4. What is the nerve that innervates the muscle derived from the third branchial arch?
   A. The trigeminal nerve
   B. Cranial nerve VII
   C. The glossopharyngeal nerve
   D. The hypoglossal nerve
   E. The recurrent laryngeal branch of the vagus nerve

Chapter 42
MORPHOPHYSIOLOGY OF THE LARYNX

Joy S. Reidenberg and Jeffrey T. Laitman

The larynx is a dynamic organ of many functions. Its critical position in the pharynx at the intersection of the nasal, oral, tracheal, and esophageal openings necessarily involves it in several systems, including the respiratory/vocal and digestive tracts. Perhaps no component of the body plays as central a role as the larynx, whose position and interactions with other structures profoundly influence such a large variety of activities.

The basic principle involved in nearly all of the laryngeal functions is the ability to regulate airflow. The larynx is actually a complex valve that evolved from a simple sphincter. This sphincter's original function was solely protection: to close off the connection of a gas bladder (which later evolved into lungs) from the digestive tract. The protective function is still one of the most important roles of the human larynx. A closed laryngeal valve combined with gross positional movements of the larynx is necessary for controlling swallowing movements and affording protection of the respiratory tract from accidental inhalation of foreign substances, particularly from the digestive tract.

A closed larynx prevents airflow and is thus also necessary for increasing intrathoracic pressure. Forced expiration against the closed laryngeal valve builds up the pressure needed for a cough or sneeze. Closure of the larynx also assists the muscular system by stabilizing the thoracic skeleton. When the volume of air trapped within the lungs is fixed, movements of the ribs are restricted. The thorax can then serve as a rigid platform for muscles used during torso or upper extremity movements. Closure of the larynx can also affect the lower digestive and urogenital regions. Maintaining stable intrathoracic pressure allows the diaphragm to be immobilized and thus rigidify the superior wall of the abdominal cavity. This wall stability enables contraction of abdominal muscles to cause a rise in intra-abdominal pressure (straining, "bearing down," or Valsalva maneuver), which helps to force contents caudally (e.g., evacuating hollow organs during defecation, forced micturation, or parturition).

The human larynx has evolved beyond its original protective function and now provides precise control over airflow. This is accomplished by regulating the diameter of the laryngeal opening and by subtle changes in the length and tension of the vibrating elements of the valve (paired vocal folds). The phonatory function of the larynx (particularly in speech production) is a later evolutionary overlay on the original protective role of the larynx, but it is one of the special traits that make humans unique among animals.

LARYNGEAL SKELETON

The larynx is located in the neck, with its long axis vertically aligned with the trachea. It articulates with the hyoid bone anterosuperiorly, is attached to the muscular walls of the pharynx dorsally, and connects with the trachea inferiorly. The laryngeal lumen is continuous with the space of the laryngopharynx superiorly and dorsally, and the lumen of the trachea inferiorly. The laryngeal skeleton is composed of three unpaired cartilages (epiglottic, thyroid, cricoid) and three paired cartilages (arytenoid, corniculate, and cuneiform) (**Figs. 42–1, 42–2**, and **42–3**). The cartilages are joined to each other, and to the hyoid bone above and the trachea below, by synovial joints, ligaments, membranes, and muscles.

The most superior (cranial) cartilage is the epiglottis, and the most inferior (caudal) cartilage is the cricoid. The ventral (anterior) aspect is largely occupied by the thyroid cartilage and also includes the epiglottis superiorly and the cricoid inferiorly. The dorsal (posterior) aspect is composed mostly of the cricoid cartilage (the thyroid cartilage is incomplete dorsally). Immediately superior to the cricoid cartilage and visible from the dorsal aspect are the paired arytenoid and corniculate cartilages. The cuneiform cartilages are also superior to the cricoid cartilage but are found slightly lateral and ventral to the arytenoid cartilages and medial to the externally overlapping thyroid cartilage. Accessory cartilages may be variably present. These are often referred to as the cartilages of Luschka and may include the triticeal (in the lateral thyrohyoid ligaments), the interarytenoid, and a small cartilage sometimes found within the vocal ligament of the vocal folds.

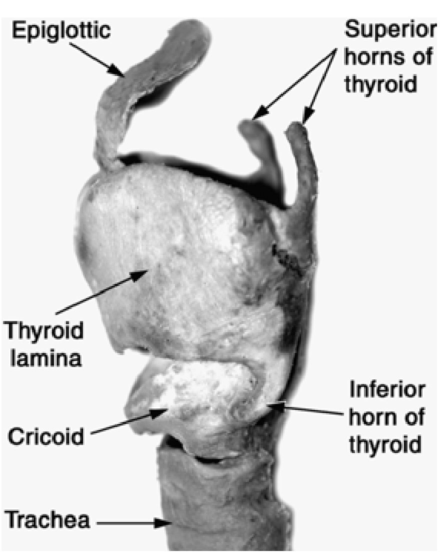


Figure 42–1 *Left lateral view of the larynx.*

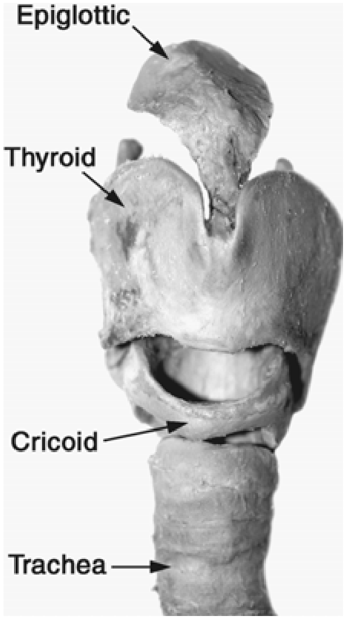


Figure 42–2 *Ventral view of the larynx.*

The epiglottic, corniculate, cuneiform, and upper portions of the arytenoid cartilages are composed of elastic fibrocartilage and thus do not ossify with age. The thyroid, cricoid, and lower portions of the arytenoid cartilages are much more rigid due to their hyaline nature and may exhibit age-related ossification.

The epiglottic cartilage is positioned in the midline and is the most superior cartilage of the larynx (**Figs. 42–1, 42–2**, and **42–3**). The cartilage and its coverings are collectively termed the epiglottis. The word *epiglottis*

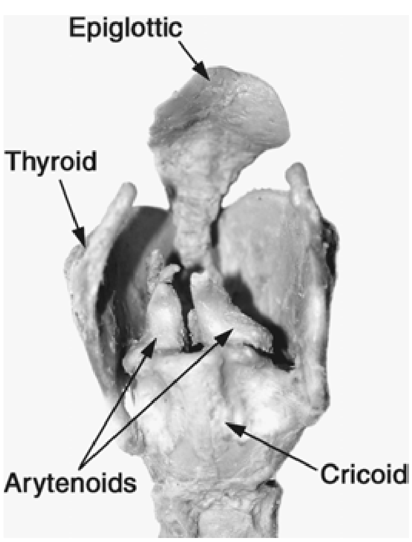


Figure 42–3 *Dorsal view of the larynx. Note that the right arytenoid cartilage is adducted and tilted posteriorly and that the cuneiform and corniculate cartilages are missing from the right arytenoid.*

indicates its position above the glottis (the level of the vocal folds). The epiglottic cartilage is highly flexible due to its high composition of elastic fibers, and its shape facilitates posterior mechanical folding. Its narrow stalk-like base is anchored to the ventral, lumenal aspect of the thyroid cartilage by the thyroepiglottic ligament. As the epiglottis extends cranially above the level of the hyoid bone, it widens to a broad "leaf" shape. It is often angled dorsally, so that its lumenal surface faces inferiorly toward the tracheal opening. The hyoepiglottic ligament (a muscle in some animals) attaches the epiglottic cartilage to the hyoid bone. The epiglottis overlaps the soft palate and projects into the nasopharynx in newborn human infants, thereby channeling airflow directly between the nasal cavity and the lungs. In this position, the epiglottis can also assist in protecting the opening to the larynx. The epiglottis obstructs food and liquid from directly entering the larynx. Its rostral surface (which faces the oral cavity) is shaped like the bow of a boat, with the midline ridge formed from the median glossoepiglottic fold (which spans between the epiglottis and the tongue). In human infants, this shape helps divert oncoming liquids and redirects the flow to the left and right lateral food channels (piriform sinuses), which pass around the laryngeal inlet and rejoin dorsally and caudally at the esophagus. Although these functions are essential in infants, the adult epiglottis may be largely vestigial due to factors relating to the lower laryngeal position in the neck (see later discussion). The adult epiglottis is thus not necessary for respiration, protection during swallowing, or phonation.

The thyroid cartilage (from the Greek *thyreos*, meaning "shield") is the largest laryngeal cartilage. It covers the ventral aspect of the laryngeal airway. The left and right sides (laminae) are joined in the ventral midline inferiorly but are cleft superiorly (thyroid notch) (**Fig. 42–2**). The superiormost point of fusion projects ventrally and is called the laryngeal prominence (Adam's apple). This protrusion is visible under the skin of the neck and is usually larger in postpubescent males. The left and right laminae of the thyroid cartilage do not unite dorsally. Rather, each lamina has a pair of projections (cornua, or horns) from the dorsal edge: one extending superiorly (superior horn) and one extending inferiorly (inferior horn) (**Fig. 42–1**). The superior edge of the thyroid cartilage attaches to the hyoid bone via the thyrohyoid membrane. The dorsal edges of this membrane are each thickened into a thyrohyal ligament. This ligament extends from the superior horn of the thyroid to the dorsal edge of the greater horn of the hyoid and often contains a small, round triticeal cartilage (tritiate cartilage of Luschka). The inferior horns of the

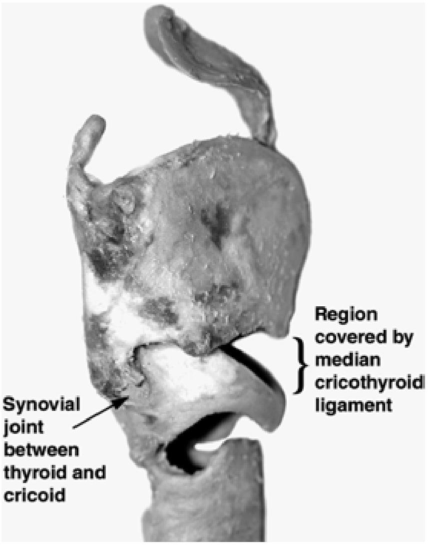


Figure 42–4 *Right lateral view of the larynx. The synovial cricothyroid joint and region covered by the median cricothyroid ligament are shown.*

thyroid cartilage attach to the dorsolateral aspects of the cricoid cartilage at synovial joints (**Fig. 42–4**). Movements at this joint allow the thyroid cartilage to rotate ventrally and dorsally about a transverse (left to right) axis relative to the cricoid cartilage (or vice versa, the joint can allow the cricoid to approximate or separate from the thyroid cartilage at their ventral aspects). The thyroid and cricoid cartilages are attached ventrally by a median cricothyroid ligament. This ligament is particularly important because it is through this structure that an emergency cricothyrotomy is performed to establish a temporary opening to relieve an obstructed airway (**Fig. 42–4**).

The cricoid cartilage (from the Greek *krikos*, meaning "ring") is a complete ring. It is wider dorsally and narrower ventrally, like a signet ring. The broad dorsal portion (lamina) extends cranially above the level of the ventral portion. This feature prevents an instrument inserted during an emergency cricothyrotomy from penetrating too deeply and making an unintended fistula to the esophagus. The cricoid cartilage is attached caudally to the first tracheal ring by a cricotracheal ligament. The cranial border of the cricoid lamina supports a pair of arytenoid cartilages. They each articulate at a synovial joint on the lateral aspects of the dorsal region of the cricoid lamina.

The paired arytenoid cartilages (from the Greek *arytaina*, meaning "ladle") are each shaped like a tetrahedron (pyramid), with the apex directed cranially and the base attaching to the cricoid cartilage. The arytenoid has a lateral extension called the muscular process and a

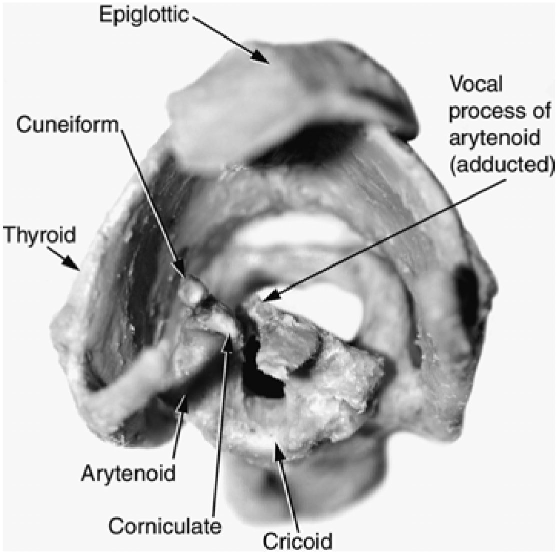


Figure 42–5 *Superior view of the larynx. Note that the right arytenoid cartilage is adducted and tilted posteriorly and that the cuneiform and corniculate cartilages are missing from the right arytenoid.*

ventral extension called the vocal process (**Fig. 42–5**). The vocal process supports the vocal ligament, which, in turn, supports the medial edge of the vocal fold (discussed later). The paired arytenoid cartilages are the most mobile structures of the larynx. The large cricoarytenoid synovial joint allows three types of movements: sliding in the dorsoventral axis, sliding in the lateral-medial axis, and rotation about the craniocaudal axis. These movements affect the positions of the vocal process, and therefore the attached vocal folds.

The corniculate and cuneiform cartilages are small, paired cartilages that border the dorsal and lateral aspects of the laryngeal aditus (opening into the lumen of the laryngeal vestibule) (**Fig. 42–6**). The corniculate cartilages (from the Latin *cornu*, meaning "horn"; also called cartilages of Santorini) lie at the cranial apex of each arytenoid cartilage. They have a cone shape with a wide base and a narrow tip that curls dorsally toward the esophageal opening. The cuneiform cartilages (from the Latin *cuneus*, meaning "wedge"; also called cartilages of Wrisberg) are positioned lateral and ventral to each arytenoid. The cuneiform has the shape of a club or mallet, being wider cranially, narrower caudally, and rounded superiorly. The corniculate and cuneiform cartilages help support the dorsal and lateral edges of the aryepiglottic folds.

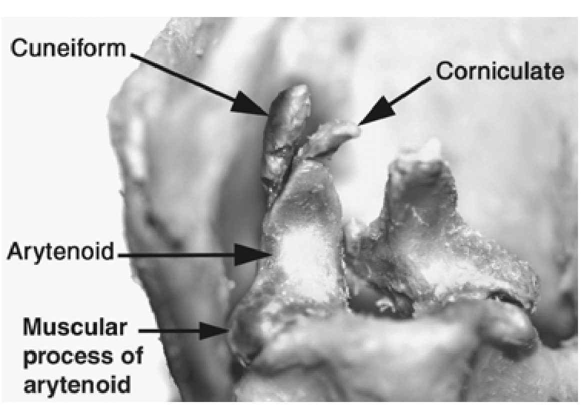


Figure 42–6 *Magnified dorsal view of the arytenoid, corniculate, and cuneiform cartilages. Note that the right arytenoid cartilage is adducted and tilted posteriorly and that the cuneiform and corniculate cartilages are missing from the right arytenoid.*

LARYNGEAL FOLDS AND MEMBRANES

The left and right aryepiglottic folds span the lateral surfaces of the arytenoid, corniculate, cuneiform, and epiglottic cartilages, whose outlines can be seen under the mucosa of the folds (**Figs. 42–7, 42–8**, and **42–9**). The superior edges of these folds define the opening into the lumen of the larynx. Lateral to each aryepiglottic fold is a piriform sinus (lateral food channel). Food or liquid passing along this pathway is prevented from

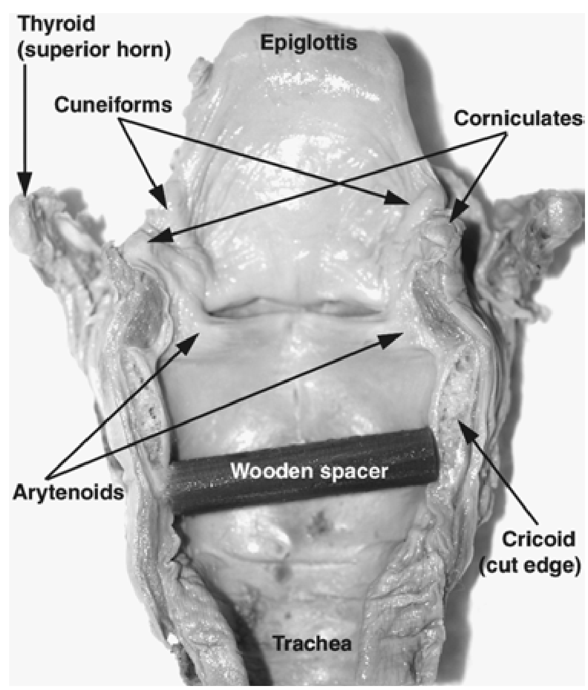


Figure 42–7 *The laryngeal lumen. The larynx is cut dorsally in the midline through the cricoid cartilage, and the cut edges are reflected laterally (held open by a wooden spacer placed in the lumen). The outlines of the laryngeal cartilages (epiglottic, cuneiform, corniculate, and arytenoid) are visible under the aryepiglottic fold. The superior horns of the thyroid cartilage are visible laterally.*

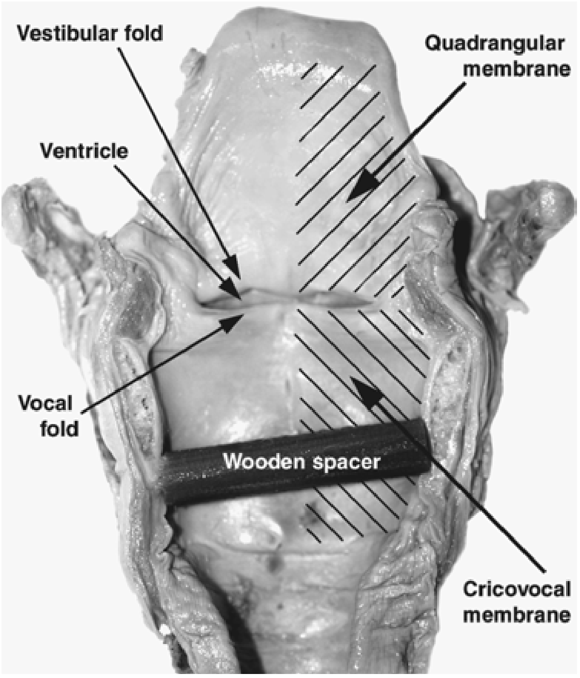


Figure 42–8 *The laryngeal lumen. As in* ***Fig. 42–7****, the larynx is cut dorsally in the midline through the cricoid cartilage, and the cut edges are reflected laterally (held open by a wooden spacer placed in the lumen). The superior to inferior relationship of the vestibular fold, laryngeal ventricle, and vocal fold is indicated on the left; the regions covered by the quadrangular and cricovocal membranes are indicated with diagonal stripes on the right.*

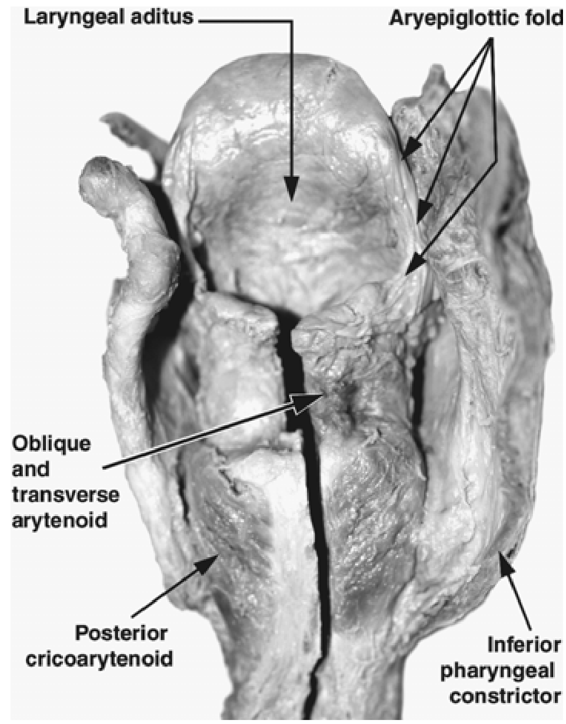


Figure 42–9 *Dorsal view of extrinsic and intrinsic laryngeal muscles. (Note: This larynx was also cut dorsally along the midline, but the cut edges remain approximated so that the laryngeal lumen is not visible.) The mucosa has been removed from the dorsal aspect, revealing the posterior cricoarytenoid, oblique arytenoid, and transverse arytenoid muscles. The cut edge of the inferior pharyngeal constrictor is also indicated. Note also the laryngeal aditus, bounded laterally by the aryepiglottic fold.*

entering the larynx by the aryepiglottic fold, which forms the medial wall of the piriform sinus and a lateral wall that protects the laryngeal aditus. Together with the cartilages, the aryepiglottic folds help form a circular cuff, with a breach in the dorsal midline called the interarytenoid notch. The low wall of this dorsal aspect is thus most vulnerable to incursions of food or liquid into the larynx, particularly from gastroesophageal reflux or regurgitation.

The interior of the larynx is lined by a fibroelastic membrane. This membrane, which runs just under the mucosa, begins superiorly as a tissue spanning the epiglottis and the arytenoid cartilages and ends inferiorly at the cricoid cartilage. The fibroelastic membrane is bilaterally symmetrical, and on each side it is divided into an upper section called the quadrangular membrane and a lower section called the cricovocal membrane (**Fig. 42–8**). The quadrangular and cricovocal membranes are separated by the laryngeal ventricle. The upper border of the quadrangular membrane supports the aryepiglottic fold.

The lower border of the quadrangular membrane terminates as a free edge, called the vestibular fold (sometimes referred to as the "false vocal cord"), which projects medially into the laryngeal lumen (**Fig. 42–8**). This fold contains the vestibular ligament, which spans dorsoventrally between the arytenoid and thyroid cartilages. The aperture between the medial edges of the paired vestibular folds is called the rima vestibuli. The function of the vestibular folds is unclear. They may assist with effort closure of the larynx, but they do not appear to contribute directly to phonation. The vestibular fold appears to have a restricted range of motion. This limited mobility may relate to its attachment proximally on the body of the arytenoid, rather than distally along any projection removed from the axis of rotation, such as the vocal process.

The free edge of the vestibular fold is formed by excavation of a pouch laterally called the laryngeal ventricle or laryngeal sinus (**Fig. 42–8**). The ventricle begins inferior to the free edge of the vestibular fold, then extends laterally. Sometimes it can also extend anteriorly and superiorly to form a large saccule under the fold. The ventricles secrete a protective layer of mucus onto the superior surfaces of the vocal folds. Their potential role in phonation as a possible resonant cavity has been suggested but remains unclear.

The cricovocal membrane contains the lateral cricothyroid ligament. This ligament spans dorsoventrally between the arytenoid and thyroid cartilages, and inferiorly along the cranial edge of the cricoid cartilage. Superiorly, it forms a thickened structure known as the vocal ligament. The vocal ligament spans dorsoventrally between the vocal process of the arytenoid cartilage and the midline, lumenal surface of the thyroid cartilage. Although the vocal ligaments may be separated or approximated dorsally (by muscular action on the arytenoid), the ventral attachments always remain fixed adjacent to each other at their connection to the thyroid cartilage. Some sources include the median cricothyroid ligament as part of the cricovocal membrane. The cricovocal membrane and its contained ligaments are sometimes collectively termed the conus elasticus. These ligaments converge medially to support a fold of tissue that projects into the laryngeal lumen known as the vocal fold (sometimes referred to as the "true vocal cord").

The vocal folds are a pair of medial projections of the laryngeal cavity walls that can be opposed in the midline to occlude the laryngeal lumen (**Fig. 42–8**). Their primary functions are in protection of the airway and regulation of airflow through the larynx. The vocal folds control inspiration/expiration, phonation, protection during swallowing, maintenance of intrathoracic/intra-abdominal pressure, and stabilization of the ribs. They are supported by the vocal process of the arytenoid cartilage dorsally and the fibroelastic cricovocal membrane containing the ligaments of the conus elasticus (the vocal ligament medially, the lateral cricothyroid ligament inferiorly, and the median cricothyroid ligament anteriorly). The vocal fold contains the vocalis muscle, which runs parallel to the lateral edge of the vocal ligament. A vocal fold appears white in color when seen from above, due to both the white color of the ligament itself and the fact that there are few blood vessels on the fold's surface. This whitish appearance contrasts sharply with the pink coloration of the vestibular fold, enabling laryngoscopic distinction. Inflamed vocal folds can appear pink, however, due to dilation of the few vessels contained within the superficial layer of the lamina propria. The aperture and the intervening space between the vocal folds is called the rima glottidis. The level of the vocal folds in the transverse plane is termed the glottis. The region above this level is referred to as supraglottic, and the region below this level is termed infraglottic or subglottic.

LARYNGEAL CAVITY

The vestibular and vocal folds divide the laryngeal cavity into three regions. The division between the upper and middle spaces is at the level of the vestibular folds, and the division between the middle and lower spaces is at the level of the vocal folds or glottis. The upper part of this cavity is called the laryngeal vestibule. The laryngeal vestibule is defined ventrally by the posterior surface of the tall epiglottis, dorsally by the low wall of the interarytenoid mucosa, and laterally by the medial surfaces of the quadrangular membrane of the aryepiglottic folds (**Fig. 42–9**). The middle part of the laryngeal cavity includes the spaces of the paired laryngeal ventricles, which are positioned between the medial projections of the paired vestibular and vocal folds. The lower part of the laryngeal cavity is called the infraglottic or subglottic space. The subglottic space extends inferiorly to the cricotracheal junction and is supported circumferentially by the circular nature of the cricoid cartilage.

The entrance into the laryngeal cavity is called the laryngeal aditus or laryngeal inlet (**Fig. 42–9**). Superiorly and dorsally, this opening connects the laryngeal vestibule with the space of the laryngopharynx. The edge of the opening is defined ventrally by the epiglottis, laterally by the aryepiglottic folds, and dorsally by the transverse mucosal fold between the arytenoid cartilages. The ventral aspect rises further superiorly than the dorsal aspect due to the height differences between the tall epiglottic and short arytenoid cartilages. Inferiorly, the subglottic lumen of the laryngeal cavity is continuous with the space of the trachea. There is a smooth transition between the lining of the cricoid cartilage and the trachea, called the cricotracheal membrane.

LARYNGEAL MUSCLES

The extrinsic muscles of the larynx are bilaterally paired muscles that are responsible for gross movements of the entire organ. Extrinsic muscles that attach directly to the larynx include the sternothyroid, thyrohyoid, and inferior pharyngeal constrictor (**Figs. 42–9, 42–10**, and **42–11**). The sternothyroid runs superiorly-inferiorly, attaching between the anterolateral aspect of the thyroid lamina and the sternum. Contraction of this muscle depresses the larynx. The thyrohyoid muscle is positioned in the same plane as the sternothyroid muscle, but it inserts superiorly and medially to it on the thyroid lamina. It extends superiorly to attach to the hyoid bone. Contraction of the thyrohyoid muscle approximates the larynx and hyoid bone. The inferior pharyngeal constrictor attaches ventrally along the lateral aspects of

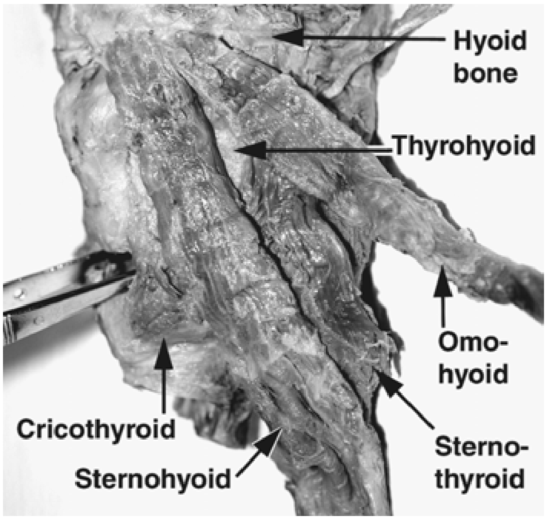


Figure 42–10 The left lateral aspect of the larynx, covered by extrinsic infrahyoid laryngeal muscles that connect to the sternum and/or hyoid. A forceps in inserted through the cricothyroid membrane, indicating the regions where an emergency cricothyrotomy would be performed.

the thyroid and cricoid cartilages, and its fibers run dorsally and superiorly to connect with fibers of the opposing side in the posterior median raphe of the pharynx. The inferior constrictor muscle is wider dorsally than it is ventrally. The upper fibers are oriented obliquely,

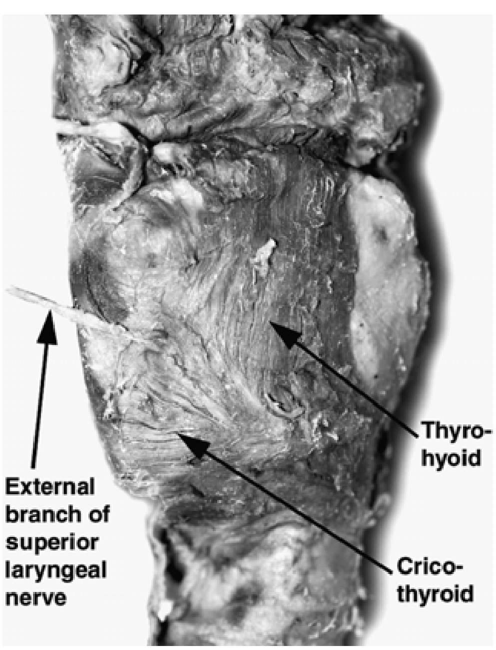


Figure 42–11 The right lateral aspect of the larynx, showing the location of the thyrohyoid and cricothyroid muscles. The external branch of the superior laryngeal nerve is seen entering the cricothyroid muscle to supply it with special visceral efferent fibers.

extending superiorly over (external to) the posterior aspect of the middle constrictor and eventually attaching via the raphe to the superior pharyngeal constrictor and then the skull base. It should be noted that this is the only direct attachment between the larynx and the skull. Contraction of the upper fibers of the inferior constrictor elevates the larynx superiorly. The lower fibers of the inferior constrictor are horizontally aligned and connect inferiorly with the esophagus. These fibers are sometimes considered to be a separate muscle called the cricopharyngeus because it appears to function independently as a sphincter that controls the esophageal opening.

There are additional pairs of extrinsic muscles that do not insert directly on the larynx but contribute to laryngeal movements by controlling movements of the attached pharyngeal walls or altering the position of the hyoid bone (which is attached superiorly to the larynx). These muscles include the middle pharyngeal constrictor, stylopharyngeus, and palatopharyngeus (which elevate the pharyngeal walls), the stylohyoid, hyoglossus, geniohyoid, mylohyoid, and digastric (which elevate the hyoid), and the omohyoid and sternohyoid (which depress the hyoid) (**Fig. 42–10**).

Several of the extrinsic laryngeal and hyoid muscles are aligned in the superior-inferior orientation and appear as long, thin bands. These so-called "strap" muscles are innervated by the ansa cervicalis (general somatic efferent fibers of C1, C2, and C3), and include the sternothyroid, sternohyoid, thyrohyoid, omohyoid, and geniohyoid. Innervations of the remaining extrinsic muscles are by the trigeminal nerve (mylohyoid and anterior belly of the digastric), facial nerve (stylohyoid and posterior belly of the digastric), glossopharyngeal nerve (stylopharyngeus), and hypoglossal nerve (hyoglossus). The palatopharyngeus and the pharyngeal constrictors are innervated by the so-called "pharyngeal plexus," which includes fibers from the glossopharyngeal, vagus, and spinal accessory nerves.

The intrinsic muscles of the larynx attach only between the cartilages of the larynx. These muscles are responsible for specific movements of the laryngeal cartilages and folds. Most intrinsic muscles act to protect the larynx by narrowing the laryngeal aditus and closing the rima glottidis (fissure between the vocal folds). This is reminiscent of the original, protective function of these muscles at an earlier evolutionary stage. Their circumferential arrangement around the laryngeal inlet suggests that they acted in concert as a sphincter to narrow the valvelike opening separating the gas bladder from the pharynx. Three muscles function to regulate the laryngeal aditus: oblique arytenoid, aryepiglottic, and thyroepiglottic. Six muscles regulate the rima glottidis: posterior cricoarytenoid, lateral cricoarytenoid, transverse arytenoid, thyroarytenoid, vocalis, and cricothyroid. All of these muscles are bilaterally paired, except for the transverse arytenoid, which spans posteriorly across the midline between the left and right sides (**Figs. 42–9, 42–10**, and **42–11**).

Only one muscle pair causes separation (abduction) of the vocal folds: the posterior cricoarytenoids (**Fig. 42–9**). These muscles originate along the dorsal laminae of the cricoid cartilage and insert on the lateral aspect of the muscular process of each arytenoid. Contractions of the posterior cricoarytenoid muscles pull the muscular processes dorsally, causing the ventrally projecting vocal processes to rotate laterally and thereby open the laryngeal lumen.

The remaining laryngeal muscles function in closure of the larynx. The majority of these act on the arytenoid cartilages, which, in turn, control movements of the aryepiglottic and vocal folds. Adduction of the vocal folds is accomplished by the lateral cricoarytenoid muscles. These muscles arise from the lateral aspects of the cricoid cartilage and insert into the lateral aspect of each arytenoid's muscular process. Lateral cricoarytenoid contraction pulls the muscular processes of the arytenoids laterally, thereby causing the vocal processes to rotate medially toward each other. The transverse arytenoid (interarytenoid) and oblique arytenoid muscles attach dorsally between the arytenoid cartilages. The transverse arytenoid is oriented in the left-right axis, and the oblique arytenoids crisscross each other in an X-shape. These muscles cause medial opposition of the vocal folds by approximating the arytenoid cartilages. The aryepiglottic muscles are aligned with the oblique arytenoid muscles so that they appear to be extensions of the obliques that travel along the aryepiglottic fold. They constrict the laryngeal inlet by adducting the aryepiglottic folds and approximating the epiglottis and the arytenoids. The thyroarytenoid muscles attach between the ventral aspect of the arytenoids and the dorsal midline of the lumenal aspect of the thyroid cartilage. These muscles relax the vocal folds by pulling the arytenoid ventrally, but they also approximate the vocal folds by rotating the arytenoids medially. The vocalis muscle is sometimes considered to be a portion of the thyroarytenoid. Each vocalis runs adjacent to the lateral aspect of the vocal ligament and functions to shorten the vocal folds (i.e., draw them ventrally). The thyroepiglottic muscles are sometimes considered to be extensions of the thyroarytenoid muscles into the aryepiglottic folds. These muscles attach between the epiglottis and the thyroid cartilage. Contraction of the thyroepiglottic muscles may depress the epiglottis, but they can act on the aryepiglottic folds to cause widening of the laryngeal inlet.

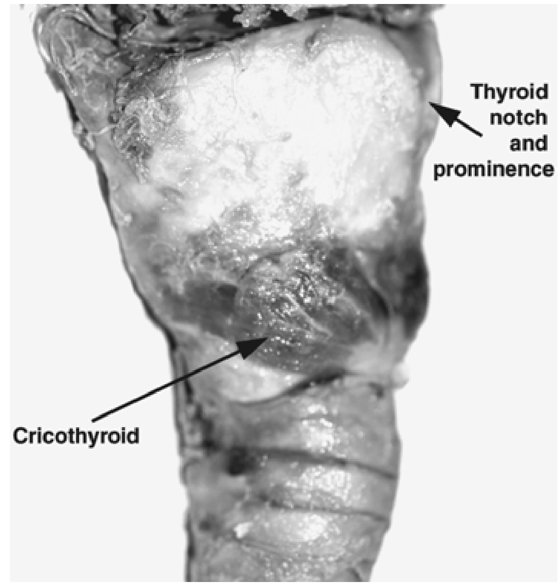


Figure 42–12 *The right lateral aspect of the larynx, with all muscles removed except the cricothyroid. The thyroid notch and prominence are indicated in this female specimen. This prominence would be more pronounced in an adult male larynx.*

The cricothyroid muscles do not attach to the arytenoid cartilages and thus can act only indirectly on the vocal folds. They are positioned along the lateral aspects of the cricoid cartilage, with the fibers running dorsoventrally (**Fig. 42–12**). Each cricothyroid has two bellies that are positioned adjacent to each other in the superior-inferior direction. Both bellies are attached ventrally near the midline of the cricoid. The superior belly attaches superiorly to the curved inferior surface of the thyroid lamina. The inferior belly attaches dorsally to the anterior aspect of the inferior horn of the thyroid cartilage. These two bellies of the cricothyroid muscle approximate the thyroid and cricoid cartilages ventrally toward each other via rotation at the synovial cricothyroid joint. This movement causes tension of the vocal folds as the connection between the arytenoid cartilages (which are attached to the cricoid) and thyroid cartilage is stretched. It is not clear whether the ventral aspect of the thyroid cartilage is pulled inferiorly, or the ventral aspect of the cricoid cartilage is pulled superiorly. The net effect is the same: the distance between the ventral attachment of the vocal folds at the thyroid cartilage and dorsal attachment to the arytenoid cartilages is lengthened.

LARYNGEAL VASCULATURE

The larynx is supplied by the superior and inferior thyroid arteries. The superior thyroid artery is a branch of the external carotid artery, and the inferior thyroid artery is a branch of the thyrocervical trunk from the subclavian artery. The larynx is drained by the superior and middle thyroid veins, which join the internal jugular vein, and the inferior thyroid vein, which empties into the left brachiocephalic vein. Lymphatic drainage is laterally to the deep cervical and paratracheal lymph nodes, and medially to the prelaryngeal and pretracheal lymph nodes.

LARYNGEAL INNERVATION

The larynx is supplied by the internal and external branches of the superior laryngeal nerve, the recurrent laryngeal nerve, and sympathetic fibers (**Fig. 42–13**). The superior laryngeal nerve is a branch of the vagus

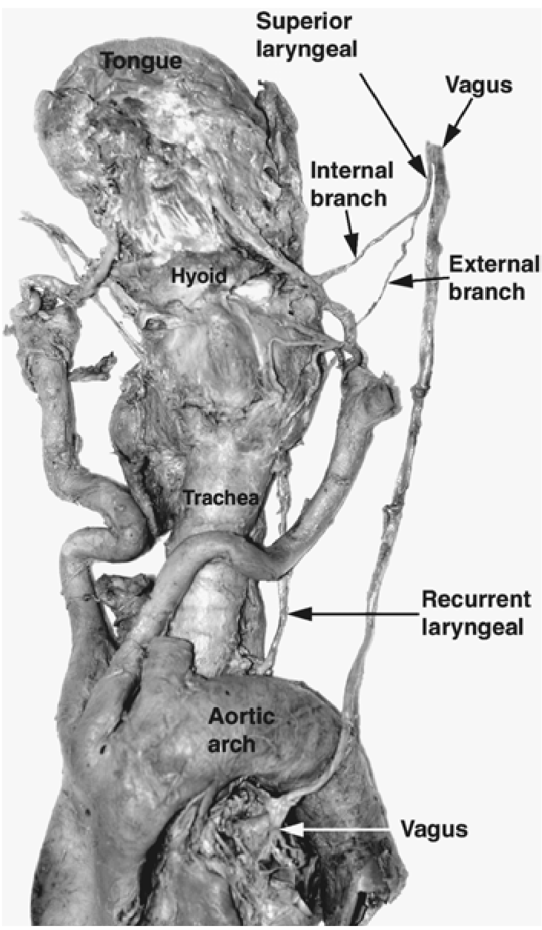


Figure 42–13 *Ventral view of a larynx with its surrounding vasculature and nervous supply. The hyoid and tongue are attached superiorly, and the trachea is attached inferiorly. Note the two branches of the vagus nerve: superior laryngeal and recurrent laryngeal. The superior laryngeal nerve divides into the internal branch (sensory) and the external branch (motor to the cricothyroid muscle). Note the path of the left vagus adjacent to the aortic arch, and the left recurrent laryngeal branch, which travels under the aortic arch adjacent to the ligamentum arteriosum. The thyroid gland has been removed on the left aspect of the larynx to reveal the path of the left recurrent laryngeal nerve lateral to the trachea.*

[cranial nerve (CN) X]. The internal laryngeal branch of the superior laryngeal nerve carries sensory innervation to and autonomic innervation from the regions above the glottis. The sensory component includes general afferent fibers to the lumen of the upper laryngeal cavity, including the superior surface of the vocal folds, and epiglottis. It also carries special visceral afferent fibers to the epiglottis for taste. The autonomic component consists of general visceral efferent fibers (preganglionic parasympathetic) from the dorsal vagal nucleus. The external laryngeal branch of the superior laryngeal nerve is a motor nerve and carries only special visceral efferent fibers to the cricothyroid muscle (**Fig. 42–11**).

The recurrent laryngeal nerve is a branch of the vagus nerve. It carries general visceral afferent (sensory) fibers from the region below the glottis, and special visceral efferent (motor) fibers from the nucleus ambiguus to all laryngeal muscles except the cricothyroid. The motor component is derived from the cranial portion of the spinal accessory nerve (CN XI). These motor fibers are carried in the vagus nerve and are distributed to the larynx by the recurrent laryngeal branch of the vagus nerve.

EMBRYOLOGY

The larynx is largely derived from the fourth through sixth branchial arches. The development likely follows a craniocaudal progression, with the more superior thyroid cartilage arising from the fourth arch, and the more caudal cricoid cartilage arising from the sixth arch. The precise arch origins of the intervening arytenoid, corniculate, and cuneiform cartilages are less clear. This is perhaps related to the absence or rapid degeneration of the fifth branchial arch during development. Unlike the other arches, the fifth arch is not clearly visualized, and thus its contributions are unknown. It is interesting to note that the glottis is the dividing line between the sensory innervation from the superior laryngeal nerve and the recurrent laryngeal nerve. Because these nerves represent innervation from the fourth arch (vagus) and the fifth-sixth arch (spinal accessory), it is possible that structures that develop in the region of the vocal folds (the territorial boundary between the fourth and sixth arches) represent what remains of the elusive fifth arch. It is thus possible that the conus elasticus, arytenoid, corniculate, and cuneiform cartilages are fifth arch elements. The epiglottis is a later mammalian development, and it does not appear to have a branchial arch origin.

LARYNGEAL POSITION

The position of the larynx varies between birth and adulthood. In newborn humans, the larynx is found very high in the neck, with the tip of the epiglottis contacting the soft palate and entering the nasopharynx. It retains this high position until ~1½ to 2 years of age. The high position allows the ventral aspect of the epiglottis to contact the soft palate and overlap its dorsal surface. This connection provides a direct pathway for channeling inspired air from the nose to the trachea. The soft palate–epiglottic connection also serves to divert food or liquid from the mouth laterally around the larynx en route to the esophagus. The high position thus enables the existence of largely separate respiratory and digestive pathways, thereby facilitating nearly simultaneous breathing almost exclusively through the nose while swallowing liquids.

The high laryngeal position limits formation of certain essential vowel sounds. Infants cannot generate the full range of vowel sounds required for adult speech because their high larynx limits the length of their vocal tract (i.e., the respiratory tract regions involved in speech production). A large supralaryngeal space (which includes the oropharynx in adult humans) is a necessary component of the vocal tract in producing speech. It acts as a filter and a resonator, and is thus largely responsible for modifying the fundamental frequencies produced by the vocal folds to generate the full range of speech sounds. Because the geometry of the vocal tract determines the vowel characteristics of speech, an elevated larynx greatly restricts this capability. In addition, a high laryngeal position is tied to the tongue being placed entirely within the oral cavity. A completely intraoral tongue cannot contribute to modifying the anterior pharyngeal wall, as the posterior portion of the tongue does in adult humans. The restricted oropharyngeal portion of the supralaryngeal space thus limits the vocal tract's ability to modify speech sounds in infants.

The high laryngeal position is not unique to human newborns and infants. It is actually a widespread pattern among terrestrial mammals. Separation of the respiratory and digestive tracts into a two-pathway arrangement allows breathing and swallowing simultaneously. This is particularly important in macrosmatic mammals because they can use the respiratory tract for olfaction (e.g., to detect the scent of a predator) while continuously feeding. Although the high position of the larynx confers respiratory protection during swallowing, it severely limits the range of sounds that can be produced. Most mammals must therefore alter the shape of the oral and nasal cavities to modify laryngeal sounds. Some species have even developed specialized air reservoirs, such as nasal or laryngeal air sacs (the latter are homologous to human laryngeal ventricles), to compensate for the restricted or nonexistent oropharyngeal space.

The larynx begins to descend after the second year of life and eventually comes to rest in the final descended position by adulthood. In this descended position, the larynx can no longer contact the soft palate or connect directly with the nasopharynx. Concomitant with this descent, the posterior portion of the tongue also descends to form the anterior wall of the pharynx above the larynx. Lowering of the larynx creates a new, permanent space between the larynx and the nasopharynx. This space, called the oropharynx, is situated just dorsal to the oral cavity. Because the larynx cannot lock into the nasopharynx, the respiratory and digestive tracts now intersect in the enlarged supralaryngeal space. Due to this crossing, it is no longer possible to breathe and swallow almost simultaneously. Thus, although adults are still habitual nose breathers, the oral connection allows breathing through the mouth.

The lowered larynx means that the respiratory tract has a greater need for protection during swallowing because food is largely passed directly over the laryngeal opening. This is accomplished initially by elevating the larynx and flexing it at the cricothyroid joint. This "folding" action tilts the arytenoid and corniculate cartilages superiorly-ventrally to approximate the epiglottis, thus largely closing off the laryngeal inlet. Simultaneously, the intrinsic muscles of the larynx approximate the vocal folds, thereby functioning as a sphincter to seal the entrance to the lower larynx and trachea. The epiglottis has become a largely vestigial structure (its main function is to enable newborns and infants to establish the direct airway between the nasopharynx and larynx). In adults, however, it can be folded over by passage of a bolus of food and thus has acquired a secondary role in protecting the laryngeal inlet during swallowing.

A low laryngeal position predisposes adults to choking due to the loss of separation between the respiratory and digestive pathways. If a bolus of food accidentally enters the laryngeal aditus during swallowing and becomes trapped there, the person may asphyxiate. Another disadvantage of the crossed pathways is the ease with which regurgitated material can be aspirated into the larynx and trachea and passed to the lungs. The lowered laryngeal position creates, in effect, a hole (the laryngeal aditus) located in the anterior pharyngeal wall just under the tongue. This low position, combined with the relatively low posterior wall of the larynx, makes the laryngeal inlet particularly vulnerable to insult from vomitus or gastroesophageal reflux.

Although laryngeal descent increases the risk of insult to the respiratory tract, it also provides an essential component for speech modification and vowel formation: a greatly expanded vocal tract. The walls of the oropharynx can be manipulated to change the shape of the vocal tract, particularly because the highly mobile posterior portion of the tongue now occupies much of the anterior pharyngeal wall. Pharyngeal modification of sounds produced at the vocal folds is thus considerably greater than that possible for newborns, infants, or any nonhuman mammal. In essence, it is the unique descent of the larynx and the resultant expansion of the pharynx that gives us the anatomical ability to produce fully articulate speech.

Suggested Readings

Harrison DFN. The anatomy and physiology of the mammalian larynx. Cambridge: Cambridge University Press; 1995

Laitman JT, Crelin ES. Developmental change in the upper respiratory system of human infants. Perinatal Neonatol 1980;4:15–22

Laitman JT, Reidenberg JS. Specializations of the human upper respiratory and upper digestive systems as seen through comparative and developmental anatomy. Dysphagia 1993;8:318–325

Laitman JT, Reidenberg JS. Comparative and developmental anatomy of laryngeal position. In: Bailey BJ, ed. Head and Neck Surgery—Otolaryngology. Vol 1. 2nd ed. Philadelphia: JB Lippincott; 1998:45–52

Laitman JT, Reidenberg JS. The human aerodigestive tract and gastroesophageal reflux: an evolutionary perspective. Am J Med 1998;103:3–11

Sasaki CT, Levine PA, Laitman JT, Crelin ES. Postnatal descent of the epiglottis in man: a preliminary report. Arch Otolaryngol 1977;103:169–171

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. Which of the following statements concerning the vocal fold is false?
   - A. It normally appears pink (the same coloration as the vestibular fold).
   - B. It is located inferior to the laryngeal ventricle.
   - C. Its fibroelastic component is the cricovocal membrane or conus elasticus, which includes the lateral circothyroid ligament.
   - D. It is supported dorsally by the vocal process, which is an extension of the arytenoid cartilage.
   - E. It contains the vocalis muscle.
2. The function of the posterior cricoarytenoid muscle is to
   - A. Adduct the vocal fold
   - B. Abduct the vocal fold
   - C. Oppose or approximate the vocal fold
   - D. Lengthen the vocal fold
   - E. Shorten the vocal fold
3. Innervation of the cricothyroid muscle is by
   - A. Special visceral efferent fibers in the recurrent laryngeal nerve
   - B. Special visceral efferent fibers in the internal branch of the superior laryngeal nerve
   - C. Special visceral efferent fibers in the external branch of the superior laryngeal nerve
   - D. General somatic efferent fibers in the vagus nerve
   - E. General somatic efferent fibers in the cranial portion of the spinal accessory nerve
4. Which of the following statements concerning the anatomy of the human newborn or infant is false?
   - A. The space of the supralaryngeal pharynx is reduced due to the high laryngeal position.
   - B. The epiglottis can contact the soft palate.
   - C. The tongue is contained entirely within the oral cavity.
   - D. The oropharynx is enlarged.
   - E. Human newborns and infants cannot produce the full range of vowel sounds necessary for speech.
5. Which of the following statements regarding the larynx is true?
   - A. It is composed of three paired (arytenoid, corniculate, and cuneiform) and two unpaired (thyroid and cricoid) cartilages.
   - B. The fundamental frequencies of phonation are generated by vibrations of the vestibular folds.
   - C. It is depressed during swallowing, and the thyroid cartilage is tilted ventrally relative to the cricoid cartilage to assist in reducing the anteroposterior diameter of the laryngeal aditus.
   - D. It is positioned high in the neck, with the epiglottis entering the nasopharynx.
   - E. It is essentially a valve and thus plays a role in regulating intrathoracic and intra-abdominal pressures and stabilizing the rib cage.

Chapter 43
NEUROLOGICAL DISORDERS OF THE LARYNX

ABIGAIL ARAD-COHEN AND ANDREW BLITZER

The larynx is composed of nine cartilages: three unpaired (the thyroid, cricoid, and epiglottis) and six paired (the arytenoids, corniculates, and cuneiforms). The hyoid bone, although technically not a part of the larynx, is also involved in its movement.

The cartilages and bony portions of the larynx are held together by mucosal folds, ligaments, and membranes. The thyrohyoid membrane is a broad fibroelastic membrane, which is referred to as a ligament. It extends from the lower border of the hyoid bone to the superior border of the thyroid cartilage, carrying the superior laryngeal vessels and nerve to the larynx. The entire larynx is lined by mucosa overlying a sheet of fibroelastic tissue. This fibroelastic membrane can be divided into two portions. The upper portion, termed the quadrangular membrane, extends between the arytenoid and epiglottic cartilages. Its superior border is free within the aryepiglottic folds, and its inferior border, the vestibular ligament, is located in the false vocal cord. The lower portion, called the triangular ligament or conus elasticus, is attached to the upper border of the arch of the cricoid cartilage. As it projects superiorly, it is attached anteriorly to the midline of the inner surface of the angle of the thyroid cartilage and posteriorly to the vocal process of the arytenoid cartilage. The anterior portion of the triangular ligament is the cricothyroid ligament. The free superior end of the ligament thickens as it extends from the thyroid cartilage to the arytenoid cartilage and forms the vocal ligament, located in the true vocal cord. The point where the true vocal ligaments meet and attach to the thyroid membrane is called Broyle's ligament.

The musculature of the larynx can be divided into an extrinsic and an intrinsic group. The extrinsic muscles elevate or depress the larynx as a whole or affect the movement of an individual cartilage. The intrinsic musculature changes the position and tension of the vocal cords. The extrinsic muscles include the infrahyoid, or strap muscles, which lie anterior to the larynx. These are the sternohyoid, sternothyroid, and omohyoid muscles. The stylohyoid, digastric, mylohyoid, stylopharyngeus, palatopharyngeus, and middle and inferior pharyngeal constrictors complete the extrinsic muscle group. The intrinsic muscles of the larynx include the cricothyroid, posterior cricoarytenoid, lateral cricoarytenoid, interarytenoid, and thyroarytenoid. The cricothyroid muscle is the only intrinsic muscle innervated by the external branch of the superior laryngeal nerve; the other intrinsic muscles are innervated by the recurrent laryngeal nerve.

Each muscle acts to tense or relax the vocal cords in a specific manner. Contraction of the cricothyroid muscle increases the distance between the arytenoid and thyroid cartilages, causing the vocal cords to be stretched and therefore tensed. The posterior cricoarytenoid muscle causes the muscular process of the arytenoid to be pulled posteriorly and toward the opposite arytenoid. The vocal process, in turn, is pulled laterally, causing the abduction of the vocal cord. The lateral cricoarytenoid muscle causes the muscular process to be pulled anteriorly and the arytenoid to rotate with the vocal process turning medially. This causes adduction of the vocal cord. The interarytenoid muscles pull the arytenoid cartilages toward one another, causing adduction of the vocal cords. Finally, the thyroarytenoid muscle alters the height and length of the vocal cord surfaces, causing the vocal cord to relax and adduct.

The vagus nerve carries the motor, sensory, and secretory (parasympathetic) fibers to the larynx. It divides into two primary branches, the superior laryngeal nerve and the recurrent laryngeal nerve. The superior laryngeal nerve originates from the vagus at the level of the inferior ganglion. The internal branch of the superior laryngeal nerve carries both sensory and secretory fibers and goes on to supply the mucosa of the epiglottis, the aryepiglottic folds, and the cavity of the larynx as far inferior as the vocal cords. The superior (motor) branch of the superior laryngeal nerve supplies the cricothyroid muscle. The recurrent laryngeal nerve, which winds posteromedially around the subclavian artery on the right and around the ligamentum arteriosum on the left, divides into anterior and posterior branches. The posterior branch supplies the posterior cricoarytenoid and the interarytenoid muscles. The anterior branch supplies the lateral cricoarytenoid, thyroarytenoid, and vocalis muscles. The mucosa of the subglottic area is also innervated via the recurrent laryngeal nerve.

LARYNGEAL SENSORY RECEPTORS

As a guardian of the airway, the larynx can initiate a wide range of reflexes with significant physiological effects. The laryngeal chemoreflexes, initiated when chemical substances contact the laryngeal mucosa, include apnea, swallowing, hypertension, and changes in peripheral vascular resistance. These reflexes can have fatal consequences. The larynx also contains a variety of mechanoreceptors in its muscles and joints, which respond to and influence laryngeal function. These mechanoreceptors can be further divided into two groups. The first group consists of touch receptors, which may act to identify foreign bodies in the airway and modulate phonation. The second group consists of the mechanoreceptors responsive to airflow, transmural pressure change, and pharyngeal receptors. These probably have a role in the reflex control of breathing. Finally, there are proprioreceptors that act as muscle spindle receptors.

LARYNGEAL MOTOR INNERVATION

The topographic representation of the laryngeal muscles was first demonstrated in the lower lateral face region of the motor cortex of the monkey. Stimulation of small areas within the motor cortex produces excitation of several laryngeal muscles, rather than a single one, and inhibition of other muscles. This response suggests that higher laryngeal motor neurons project to lower motor neurons in different areas of the nucleus ambiguus (NA). The NA is defined as the nuclear group containing the laryngeal lower motor neuron (LMN) and probably includes a subdivision referred to as the retrofacial nucleus (RFN). The adductor LMN is located in the dorsolateral division in the caudal two thirds of the NA, and it projects over the recurrent laryngeal nerve. The LMN for the PCA is located in the middle one third of the NA but is ventral to the adductor neurons. This relationship may have functional significance. Stimulation of the laryngeal motor cortex produces bilateral contraction of the adductor and abductor muscles. This pathway is polysynaptic, rather than a direct projection from the cortex to the laryngeal motor neurons in the NA. Other cortical and subcortical areas have a role in involuntary phonation such as in various emotional states (fear, anger, rage, and anxiety). Stimulation of these subcortical areas evokes nonverbal responses, whereas ablation of their midbrain projection sites in the periaqueductal gray units leads to a total loss of phonatory ability (mutism).

An understanding of laryngeal innervation provided by anatomical and physiological studies provides a basis for the diagnosis of hyperactive and hypoactive (paretic or bradykinetic) voice disorders. Abnormal vocalization noted in microstimulation studies of periaqueductal gray units in the monkey suggests that pathology in this center may contribute to voice disorders such as spasmodic dysphonia.

LARYNGEAL ELECTROMYOGRAPHY

To evaluate neuromuscular disorders affecting the larynx, electrodiagnosis may be valuable in making distinctions between neuropathy, anterior horn cell disease, brainstem lesion, myopathy, and neuromuscular transmission disorders. Electromyography (EMG) also may be useful in determining whether an immobile cord is related to paralysis, synkinesis, or mechanical limitation.

Electromyography may be used to document and evaluate tremor disorders, tics, and myoclonic jerks. In some hyperfunctional disorders, such as laryngeal dystonia, abnormalities of motor unit form result in particularly high-amplitude readings. In other studies, the essentially normal findings are important to exclude various conditions.

To investigate the two components of vagal innervation, percutaneous laryngeal EMG recordings of both the cricothyroid and thyroarytenoid (vocalis) muscles are performed. Because EMG activity depends on the activity of muscles, it is necessary to record spontaneous potentials at complete rest and motor unit potentials (MUPs) activated by phonation. The posterior cricoarytenoid muscle also can be studied because it is the only abductor of the larynx with motor unit activation on deep inspiration or sniffing. Studying the MUPs in the PCA is important in patients with stridor and breathy dysphonias.

HYPOFUNCTIONAL DISORDERS

Paralysis and Paresis of Vocal Folds

Situated at the crossroads of the airway and food passages, the larynx must provide three sometimes disparate functions: airway support, airway protection, and voice. Weakness or immobility of one or both vocal folds impacts all of these functions. Adequate evaluation requires careful history and physical examination, including laryngoscopy. Diagnostic accuracy permits the physician to intervene appropriately and at the right time. Laryngeal EMG allows differentiation of etiologies of vocal fold immobility, including mechanical fixation, paresis, synkinesis, and paralysis. Any or all of the nerves that subtend the laryngeal function can be affected by injury or disease. Isolated sensory loss is unusual but can occur in the absence of motor deficit. Guillain-Barré syndrome, diabetes, and "idiopathic" superior laryngeal nerve paralysis are among possible causes. Throat clearing, paroxysmal coughing, and vague foreign body sensation can be seen in unilateral loss. Fortunately, bilateral sensory loss is uncommon because it often leads to severe aspiration and pneumonia.

Superior laryngeal nerve paralysis most often combines sensory loss with weakness or paralysis of the cricothyroid muscle. It usually presents as an isolated phenomenon and is thought to be caused by viral neuropathy. Diplophonia, easy fatigability of the voice, and rotation of the posterior commissure toward the side of the paralysis during phonation are common. Approximately 60% of patients recover spontaneously within 1 year of onset. Many others compensate to a reasonable degree. Presently, there is no medical or surgical treatment available. Nerve–muscle pedicle reinnervation has been suggested by some authors.

The etiologies of recurrent laryngeal nerve paralysis include thyroidectomy or other trauma, neurological disorders, malignancy, and miscellaneous. Unilateral paralysis results in voice change. Typically, the voice becomes weak, breathy, and often diplophonic. The paralyzed vocal fold rests in the paramedian position due to the unopposed medializing pull of the intact cricothyroid muscles. The patient may complain of shortness of breath while speaking due to rapid air escape. Bilateral vocal fold paralysis is often a result of surgery or other trauma. Because the intact cricothyroid muscles adduct the cords close to the midline, voice is usually close to normal, but pitch control is poor. Airway compromise results from an inability to abduct the vocal folds, although some patients can get by with the restricted airway and can be misdiagnosed as having asthma or chronic bronchitis.

A combination of paralysis of both the recurrent and superior laryngeal nerves is less common and is usually of central origin or a result of injuries near the base of the skull. In cases of superior and recurrent nerve unilateral paralysis, the vocal cords often rest in the intermediate position, and the voice is generally very weak and breathy and does not compensate well. Unexplained paroxysmal cough or even aspiration can occur because of the loss of sensation and the glottic incompetence.

Bilateral superior and recurrent laryngeal nerve paralysis is very uncommon and results in airway compromise, poor voice, and severe aspiration.

Incomplete paralyses or palsies can occur and are often responsible for swallowing and voice complaints. The laryngeal findings, if any, are usually subtle. The recurrent laryngeal nerve also innervates the cricopharyngeus muscle; hence, a palsy can present as dysphagia due to delayed opening of the superior esophageal sphincter.

Management

Spontaneous recovery often occurs within 1 year from onset. Moreover, satisfactory compensation is common in uncomplicated unilateral recurrent laryngeal nerve paralysis. Accurate diagnosis is critical to produce a logical treatment plan that does not prevent recovery or compensation.

Unilateral sensory loss usually does not require any intervention. Severe aspiration due to bilateral sensory loss should be treated appropriately. Speech therapy is probably appropriate in noncompensated cases of superior laryngeal nerve paralysis with poor voicing.

All patients with unilateral recurrent laryngeal paralysis should be referred for evaluation and recommended for speech therapy. One year should generally be allowed for recovery or compensation to satisfactory voice levels. Several surgical approaches are available for patients who have not recovered.

Teflon or Gelfoam Injections

This is the procedure of choice for most patients. It is usually performed under topical anesthesia as an outpatient. Gelfoam is absorbed in ~2 to 3 months and can be used as a temporary treatment even for patients who still have hope of recovery. Teflon is a permanent material that can be removed, but with difficulty. These materials should be injected as far posteriorly and lateral to the paralyzed vocal folds as possible. Injections into the membranous vocal fold itself should be done infrequently and judiciously because Teflon can cause fibrosis and change the mass characteristics and therefore the vibratory qualities of the vocal fold. Complications of Teflon injection include extrusion or displacement, granuloma formation, overinjection (airway compromise), and unsatisfactory voice results.

Surgical Medialization

When the arytenoid is fixed as well as paralyzed, or when there is a large posterior commissure gap, Teflon may not be adequate to correct vocal fold incompetence. Various techniques exist to implant cartilage, muscle, or alloplastic materials surgically lateral to the immobile fold to displace it toward the midline. It is performed under local anesthesia without the need for tracheotomy. The implant is inserted on the side of the paralyzed fold through a window in the thyroid ala in the subperichondrial plane, medializing and supporting the fold. Although this is an open surgical procedure, it can overcome almost any size defect and provides access for mobilization of a fixed vocal fold. Positioning can be monitored fiberoptically, permitting precise placement. In some cases, an arytenoid adduction procedure is also needed to position the arytenoid correctly and close the posterior glottal gap.

Reinnervation

One of several techniques in which the adductors and tensors of the vocal fold can be reinnervated is by using a nerve–muscle pedicle derived from the anterior belly of the omohyoid muscle. This procedure is feasible only if the vocal fold is not fixed. The pedicle is transposed to the fibers of the lateral thyroarytenoid muscle exposed through a window in the thyroid cartilage. It can be combined with a medialization procedure as described previously to achieve an immediate improvement.

Bilateral vocal fold paralysis usually requires innervation to secure the airway. The three general approaches include tracheotomy, posterior cordotomy, and surgical lateralization.

Tracheotomy

This is almost always necessary as an initial procedure. It solves the airway problem without decreasing the voice. A "permanent" flap-type tracheotomy can be a good long-term solution for patients who cannot undergo reinnervation due to fixation of the vocal cord because it preserves residual voice.

Posterior Cordotomy

This endoscopic procedure uses a laser to cut the vocal cord medially to laterally at the attachment of the vocalis process. The muscle retracts anteriorly, leaving a keyhole opening, which increases the size of the airway but decreases the quality of the voice.

Surgical Lateralization

Many techniques exist to remove the arytenoid or lateralize the paralyzed vocal fold. All of these gain airway space at the expense of the residual voice, although tracheotomy can be dispensed if the procedure is successful.

Nerve–Muscle Pedicle Reinnervation

If the vocal folds are passively mobile, and the nerve pedicles have not been compromised by surgery, radiation, or the underlying disease, reinnervation can restore abduction without compromising the voice. Nerve–muscle pedicles can be transposed to the PCA muscles to achieve eventual (2–6 months) spontaneous abduction during respiratory effort.

Patients with combined superior and recurrent laryngeal nerve palsies are less likely to benefit from Teflon injection due to the more abducted position of the vocal fold. Surgical medialization or reinnervation is recommended for these cases to obtain good voice. Bilateral combined paralysis may be life threatening, not because of inadequate airway, but because of the inability to protect the airway due to sensory loss and an open glottis. Tracheotomy is required to secure the airway and for tracheobronchial toilette in mild or intermittent aspiration. When severe aspiration occurs, feeding by gastrostomy tube or esophagostomy should be considered. If the patient aspirates his or her own secretions, procedures such as laryngeal closure, laryngotracheal diversion, and even laryngectomy should be considered.

When dysphagia due to cricopharyngeus incompetence occurs as a result of incomplete paralysis, cricopharyngeal myotomy or injections of botulinum toxin (Botox) to the cricopharyngeus can improve swallowing.

MOVEMENT DISORDERS

Patients are classified as having movement disorder of the larynx if they have a disorder of motor programming resulting in either a paucity of movement (bradykinesia or akinesia), as in Parkinson's disease or chorea, excessive movement (hyperkinesia), as in essential tumor, spasmodic dysphonia, stuttering, myoclonus, tics, and tardive dyskinesia, or a combination of both. Proper diagnosis and treatment require a team approach, including an otolaryngologist, a neurologist, and a speech–language pathologist.

Parkinsonism

Parkinsonism is a neurological syndrome manifested by any combination of tremor at rest, rigidity, bradykinesia, and loss of postural reflexes. In Parkinson's disease, speech production is compromised due to hypokinetic dysarthria with poor presentation or air to the vocal apparatus. A reduced range of articulation occurs for both lingual and labial sounds. Laryngoscopy often reveals bowing of the vocal folds with midcord opening of the glottis on phonation. The vocal fold motion is often reduced, and there may be pooling of secretions in the hypopharynx. Therapy usually improves speech, but the pharmacological treatment can sometimes worsen speech by making it more rapid and less distinct, with some "freezing" of speech. The Silverman technique was developed to treat patients with Parkinson's disease. It involves an intensive treatment for several weeks, training patients to speak loudly. Vocal fold augmentation can be used for some patients with significant bowing of the vocal folds.

Dystonia

Dystonia is a syndrome dominated by sustained muscle contractions, frequently causing twisting and repetitive movements or abnormal postures that may be sustained or intermittent. Spasmodic dysphonia (SD) and laryngeal dysphonia are terms used to describe an action-induced laryngeal motion disorder. In SD patients, the vocal apparatus is usually normal at rest but functions abnormally with speaking. In the adductor type, there is an abnormal co-contraction of the vocalis muscle complex, resulting in inappropriate adduction of the vocal folds. These patients have a strained-strangled voice that is harsh, often with tremor, inappropriate pitch or pitch breaks, breathiness, and glottal fry. Abductor SD is characterized by inappropriate co-contraction of the posterior cricoarytenoid muscles during the action of speaking, resulting in a breathy, effortful, hypophonic voice with abrupt termination of voicing, causing whispered segments. Respiratory adductor laryngeal dystonia is a rare entity in which the involuntary adduction of the vocal folds occurs during inspiration, leading to stridor. Although systemic pharmacotherapy, speech therapy, and surgery have been used in the management of SD, intermittent Botox injections have become the predominant method of controlling symptoms. Botox is injected into the vocalis muscles bilaterally in adductor SD patients and into the PCA muscle unilaterally or staged bilaterally in abductor SD patients. Injections are performed on an outpatient basis every 3 or 4 months or when the patient feels a recurrence of symptoms. Complications include weakness of the adductors (leading to breathy voice that may last up to a month), bleeding, swelling of the vocal folds, and temporary swallowing difficulty. Abductors should not be injected bilaterally at the same time to avoid airway compromise due to inability to abduct the cords.

Tremor

Tremor has been described as a series of involuntary, relatively rhythmic, purposeless oscillatory movements. The proposed underlying neural basis for tremor includes central and peripheral mechanisms. Normal physiological tremor has a frequency of 6 to 12 Hz, whereas pathological tremor has been reported to range from 3 to 7 Hz and is considered a sign of neurological disorder. Essential tremor (ET) is a rhythmic, oscillatory movement of 4-12 Hz with variable amplitude. Vocal tremor occurs in 10 to 20% of patients with ET. The amplitude of tremor is greater with emotional stress or fatigue. Pitch breaks and phonation arrests have been associated with visible vertical oscillations of the larynx.

Vocal tremor has been observed in 30% of patients with SD, and the tremor is usually irregular. The effect of medication on reducing vocal tremor is equivocal. Botox injections to the involved muscles may show a dramatic benefit. The tremor may not be totally eliminated, but the amplitude diminishes, and voicing is more fluent. A small percentage of SD patients have tremulousness of speech in the same frequency as the tremor of their lips and jaw. Irregular vocal tremor is observed in patients with cerebellar ataxia. Rapid tremor has been observed in myasthenia gravis patients.

Stuttering

Stuttering is a neurological disorder in the tic disorder category. It includes abnormal involuntary and inappropriate use of the muscles of speech production, resulting in dysfluency. There is increased muscle tension during speech, resulting in postures that are sustained for longer than expected or in quick repetitive movements of the same posture.

Myoclonus

Myoclonus refers to sudden, brief, ticlike involuntary movements caused by muscular contractions (positive myoclonus) or inhibitions (negative myoclonus) arising from the CNS. Laryngeal involvement can become problematic when there are myoclonic jerks of laryngeal muscles and muscles of respirations. The myoclonic jerks may cause dysphonia, dysphagia, or aspiration. Singulus (hiccup) and palatal myoclonus are of particular importance to the otolaryngologist. Eyes, face, palate, larynx, diaphragm, neck, shoulder, and arm can be involved at the same time, giving rise to the syndrome of *myoclonies velopharyngolaryngooculodiaphragmatique*. Laryngeal involvement may produce a broken speech pattern, simulating that heard in SD or tremor. Examination of the vocal folds often shows the rhythmical adduction and abduction of the vocal folds at the same timing and frequency as the other involved areas. Various pharmacological and surgical treatments have had varying degrees of success in the treatment of myoclonus. Botox injection of the palate or larynx may reduce the symptom by reducing the amplitude of the myoclonic jerk.

NEUROMUSCULAR DISORDERS OF THE LARYNX

Medullary Disorders

Motor neuron disease, postpolio syndrome, syringomyelia, Arnold-Chiari malformation, and medullary brainstem strokes can result in laryngeal paralysis. Motor neuron diseases comprise a group of disorders characterized by progressive lower motor neuron signs including weakness, wasting, and fasciculation, often with bilateral corticospinal tract signs. They include progressive spinal muscular atrophy (PMSA), progressive bulbar palsy (PBP), and amyotrophic lateral sclerosis (ALS). Acute poliomyelitis may cause acute lower motor neuron paralysis, but it is now rare in industrialized countries.

Lower motor neuron disorders may include hypophonia, hoarseness, dysphagia, and dysarthria combined with specific other signs of the disorder. The postpolio syndrome results in vocal cord paralysis, which may occur years after recovery from acute poliomyelitis. Laryngeal weakness and loss of pharyngeal reflexes may be seen in syringomyelia. The Arnold-Chiari malformation in association with myelomeningocele was considered a cause of progressive choking, apnea, and aspiration in infants. Laryngeal weakness may follow medullary brainstem infarction in the territory of the posteroinferior cerebellar artery. Symptoms include dysphonia, dysphagia, loss of pharyngeal reflexes, and paralysis of vocal cords in combination with ipsilateral loss of pain and temperature on the face and contralateral loss on the trunk and extremities, vertigo, nausea, vomiting, and ipsilateral Horner's syndrome.

Injury to Laryngeal Nerves

Generally, paralysis of peripheral origin can be divided into that due to a lesion at or central to the nodose ganglion and that where the lesion is more distal. In proximal cases, all of the nerves to half of the larynx tend to be involved. Additionally, other cranial nerves are frequently affected. Disorders of the posterior cranial fossa, including meningioma, meningitis, and trauma, may affect several cranial nerves. If the lesion is near the jugular foramen, cranial nerves (CN) IX, X, and XI are affected. General clues to the presence of laryngeal paralysis in association with paralysis of CN IX to XII are always an indication for careful investigation of the upper neck, base of the skull, and posterior pharyngeal fossa, as well as examination of the ear and nasopharynx.

Tuberculosis and sarcoidosis may cause laryngeal neuropathy. There is usually evidence of lesions in other organs. Guillain-Barré syndrome is an acute demyelinating disease of peripheral nerves that usually follows viral infection or immunization. Patients with oropharyngeal, laryngeal, and respiratory weakness are at risk for aspiration. Pharyngeal nerves may be involved in generalized neuropathy, as seen in diabetes, idiopathic neuropathy, rheumatoid arthritis, polyarteritis nodosa, lupus erythematosus, chronic alcoholism, and drug toxicity. Acute porphyria may cause vocal cord paralysis in addition to severe limb weakness. The neurotoxicity of vinca alkaloids, vincristine, and vinblastine is recognized, and cases of laryngeal nerve paralysis have been reported.

Local tumor or metastases may compress the recurrent laryngeal and vagus nerves. Radiation-induced paralysis of the laryngeal nerve should be considered. Mediastinal and lung tumors, as well as cardiovascular lesions, commonly affect the left recurrent laryngeal nerve. Benign thyroid masses at the correct anatomical site can cause injury to the nerve. The neurological complications of carcinoid tumor can cause recurrent laryngeal nerve paralysis. Traumatic and surgical lesions of laryngeal nerves have been discussed previously. Idiopathic paralysis accounts for 5 to 10% of cases; viral etiology is suspected as a possible cause.

Neuromuscular Junction Disorders

Myasthenia gravis, an autoimmune disorder of the neuromuscular junction, is characterized by fluctuating ocular or oropharyngeal weakness, often with limb weakness, which improves after administration of cholinergic drugs. Drugs including quinidine and phenytoin may produce myasthenic syndrome or unmask or exacerbate a preexisting disorder of neuromuscular transmission. Acute botulism and Lambert-Eaton syndrome are neuromuscular disorders caused by presynaptic blockade. All of the neuromuscular junction disorders may cause severe weakness of cranial muscles with risk of laryngeal weakness, stridor, and aspiration. An EMG with decreasing amplitude with repetitive tasks is suggestive of a neuromuscular junction disorder. A Tensilon test and an angiotensin converting enzyme (ACE) antibody test are diagnostic of myasthenia gravis.

Disorders of Muscle

If muscles of the oropharynx or larynx are involved in a myopathic disorder, there is a risk of aspiration. Polymyositis and dermatomyositis can cause dysphagia, aspiration, and respiratory complications. The muscular dystrophies are characterized by inherited, progressive weakness with variable age at onset, distribution, and disability. Oropharyngeal and/or laryngeal weakness may be present. The metabolic myopathies, defined by known biochemical defects, can present with severe generalized weakness, infantile hypotonia, and respiratory difficulties. These include acid maltase deficiency, brancher enzyme deficiency, and cytochrome-*c* oxidase deficiency. Episodic muscular weakness severe enough to cause complete paralysis can occur in a disorder termed periodic paralysis. It is due to excessive fluctuations of total body potassium, often precipitated by exogenous factors, including diet, cold, medication, and intercurrent infection.

LARYNGEAL DISORDERS IN STROKE

The larynx and pharynx are often involved in the dysfunction seen after cerebrovascular accident. It has been estimated that 10% of all cases of vocal cord paralysis are related to a central etiology including stroke, tumors, trauma, and infection. Dysphonia from laryngeal paresis or paralysis or a movement disorder should not be confused with aphasia, dysarthria, or dyspraxia. The latter symptoms may accompany laryngeal dysfunction or occur independently.

As mentioned earlier in this chapter, a primary dysfunction of the larynx will be caused by damage to the nucleus ambiguus, which provides ipsilateral special visceral efferent fibers to the pharynx, larynx, and cervical esophagus, and the nucleus solitarius, which provides special visceral afferents and the general visceral afferents. These vagal nuclei are laterally positioned in the brainstem and receive their blood supply primarily from the posteroinferior cerebellar artery (PICA). When the damage is supranuclear, there may be dysphonia without vocal cord paralysis. The dysphonia should be put in context with the other symptoms, which include lack of voluntary speech, a hypoglossal weakness, and, often, hemiplegia. The cough reflex is most often intact. The voluntary oral phase of swallowing is compromised, but the involuntary phases are intact.

When the damage is nuclear, the dysphonia may be associated with a definitive cord paralysis. The damage is usually related to occlusion of the PICA or vertebral artery, producing a lateral medullary infarction (Wallenberg's syndrome). There is usually severe dysphagia with poor function in the pharyngeal and esophageal phases of swallowing and often a relative cricopharyngeal achalasia. In addition, the patient may have a paralyzed palate, Horner's syndrome, facial numbness, vertigo, and nystagmus. Hiccups may occur, related to the dysfunction of the respiratory center or vagal fibers. Headaches and tachycardia may occur related to damage to CN V and vagal nuclei, respectively. These laryngeal findings should be differentiated from a laryngeal dysfunction of peripheral nerve origin. Severe dysphagia may be a prominent symptom in cases of pure motor stroke as part of severe hemiparesis or hemiplegia. Some of the dysphonia seen in stroke is related to disordered motion rather than paresis. The syndrome of palatal myoclonus is associated with a dysphonia and dysphagia. Speech and swallowing may be extremely affected by oral-buccal-lingual dyspraxia. This syndrome arises from damage in the inferior frontal region of the premotor cortex.

Laryngeal paralysis with sensory deprivation secondary to vagal nerve dysfunction is very serious. If the patient is unable to keep the airway clear, intubation or tracheostomy is imperative to prevent life-threatening soilage of the airway. Feeding can be provided with a nasogastric tube or gastrostomy. Medialization of the vocal cord may help vocal function when there is an open glottic chink. EMG can be helpful in assessing the likelihood of neural return and help in planning the best treatment for the specific patient.

Suggested Readings

Blitzer A et al, eds. Neurologic Disorders of the Larynx. New York: Thieme Medical Publishers; 1992

Gluckman J, ed. Renewal of Certification Study Guide in Otolaryngology. American Academy of Otolaryngology–Head and Neck Surgery; 1998

Stemple JC, ed. Voice Therapy: Clinical Studies. San Diego: Singular; 2000

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. The etiology of dysphonia in Parkinson's disease is
   - A. Bowed vocal cords
   - B. Poor air presentation to the larynx
   - C. Vocal tremor
   - D. A and C
   - E. B and C
   - F. All of the above
2. Spasmodic dysphonia is
   - A. A focal dystonia
   - B. A form of essential tremor
   - C. A psychogenic disorder
   - D. All of the above
   - E. None of the above
3. Botulinum toxin works on hyperfunctional disorders by
   - A. Increasing the amount of acetylcholine at the neuromuscular junction
   - B. Decreasing the amount of acetylcholine in the neuromuscular junction
   - C. Stimulating norepinephrine production
   - D. Decreasing the dopa in the brain
   - E. None of the above
4. Laryngeal myasthenia gravis is often accompanied with
   - A. Eyelid droop
   - B. Easy fatigue on motion
   - C. Serum antibody to acetylcholine
   - D. Thymoma
   - E. All of the above
   - F. None of the above

Chapter 44

BASICS OF VOICE PRODUCTION

John S. Rubin and Ronald C. Scherer

This chapter covers certain basics of voice production, of both the normal and the pathologically altered larynx. Biomechanics of the larynx is emphasized on anatomical and physiological bases. Subglottal and intraglottal pressure, glottal flow, vocal fold oscillation, voicing onset and cessation, and fundamental frequency are reviewed in the normal larynx from this perspective.

Abnormal voicing is reviewed initially through an analysis of glottal skewing, phonatory instabilities, and breathy and pressed phonation. Some aspects of benign mucosal disease are then reviewed from the standpoint of the effects of increased mass and stiffness of the vocal folds in differing pathological conditions, and the clinical and acoustical correlates.

OVERVIEW

The larynx is an intrusion into the pharynx that subsumes many functions, including airway protection, pressure valving, and phonation. This chapter deals predominantly with the latter. It must be remembered, however, that airway protection is the larynx's most critical function from a survival standpoint. It is a complex sphincter designed to protect the lower airway from food particles, fluids, and saliva. It does so passively and actively; passively by diverting food around the larynx and into the piriform sinuses, actively through a series of coordinated muscular activities occurring on a subconscious level with each swallow, resulting in a closed, protected sphincter. The intrinsic muscles act predominantly in this capacity, depressing and tilting the epiglottis, medially compressing and adducting the true and ventricular vocal folds, with each swallow.

Pressure valving also has phylogenetic survival function. In this capacity, the larynx closes off the airway to prevent ingress or egress of air, thereby allowing sudden increases in intrathoracic and intra-abdominal pressures to occur. This permits activities such as forceful micturition, defecation, and weight lifting.

Phonation is arguably important for survival. It can be considered, phylogenetically, to be a more recent activity. The capacity to produce highly complex phonatory behavior appears to be limited to human beings. Laitman and Reidenberg (1993) related phonation to the lowering of the laryngeal complex from the basicranium. Its function is primarily subsumed by three groups of muscles: the intrinsic laryngeal muscles related to the arytenoid cartilages, the medial portion (vocalis) of the thyroarytenoid muscles, and the cricothyroid muscles.

As a complex valving system, the larynx operates simplistically at two time scales, one gross, one fine. The gross time scale refers to the relatively slow motion of the arytenoid cartilages, which, to varying degrees, separate to open up the airway or come together to close (or nearly close) the airway. These motions are significantly involved in the activities of breathing, blowing, sucking, yawning, initiating or ceasing voicing, voiceless consonant production, musical instrument playing, coughing, throat clearing, swallowing, staccato whistling, whispering, and effortful behaviors such as lifting and defecation. The fine time scale refers to the motion of the vocal folds during phonation; this takes place when the vocal folds are close enough, the transglottal pressure high enough, and the vocal tissues healthy enough (relative to stiffness, mass, viscosity, and contour) to permit oscillation to occur. Fundamental frequencies range from ~80 Hz in very low-pitched male voices to over 1000 Hz in very high-pitched infant crying and soprano singing. Phonation occurs for all voiced vowels and voiced consonants during speech, as well as nonspeech (but highly communicative) voiced sounds such as grunts, voiced sighs, and whining. Normal speech involves the ongoing sequencing of the gross and fine motions, from the open glottis conditions for voiceless sounds (e.g., */s, t, f/*) to closed glottis conditions for voiced vowels and consonants (e.g., */a, I, z, d, v/*).

BIOMECHANICS OF THE LARYNX

The larynx is a complex structure consisting of cartilages, muscles, ligaments, joints, and mucous membranes, all interacting in highly precise and timed functions. Given the sphincteric nature of laryngeal closure, it is the view of these authors that it is overly simplistic to argue that any given phonatory behavior occurs on the basis of a single muscular activity (although this has been, and will undoubtedly continue to be, argued so for years, for example, that the posterior cricoarytenoid muscle is the sole abductor of the larynx). The relative pull of each muscle is balanced by the activities of other muscles. The overall vector of pull of antagonistic muscle groups and joint axes determines the actual motion. This is particularly apt when discussing the various muscles (interarytenoid, lateral cricoarytenoid, etc.) that have direct attachment to the muscular process of the arytenoid cartilages and are involved in medial positioning of the vocal folds.

The underlying joint supporting this movement, the cricoarytenoid joint, is basic to phonatory control, as is the cricothyroid joint. These joints will be discussed in some detail because they are the platform on a biomechanical level from which all follows.

The cricothyroid joint is plane synovial in type. The articular facets of the cricoid cartilage, which sit laterally near the junction of the arch and the body of the cricoid, articulate with corresponding facets of the inferior horns of the thyroid cartilage. The facets face dorsolaterally and slightly superiorly. They are often grossly asymmetric, one side to the other. A joint capsule and two ligaments stabilize the joint. The posterior ligament prevents spreading of the inferior horns of the thyroid to the lateral ligament limits but does not abolish posterior displacement of the thyroid over the cricoid.

Rotation about the transverse axis is possible with opening and closing of the cricoid arch in relation to the lower border of the thyroid cartilage. This movement is predominantly controlled by contraction of the vertical belly of the cricothyroid muscle. Upon contraction it approximates the cricoid arch to the anteroinferior thyroid cartilage. In doing so it causes a relative increase in the distance between the posterior cricoid (and thereby the arytenoids) and the anterior thyroid cartilage, lengthening the vocal folds and placing them under stretch. Dickson and Maue-Dickson (1982) identified stretch of 25% in the length of the vocal ligament in fresh cadavers by this movement. Vocally, this may cause a rise in the fundamental frequency and its correlate, pitch (see later discussion).

The cricoarytenoid joint is a synovial load-bearing joint. The posterior aspect of the upper cricoid has two elliptical facets, each measuring ~6 mm in adult males along the length of their major axis. Each facet slopes laterally, downward, and forward. The base of each arytenoid cartilage, on the inferior surface of its muscular process, has a corresponding facet, which articulates with it.

There are two ligaments and a tight fibrous articular capsule that affect the motion of the cricoarytenoid joint. The posterior cricoarytenoid ligament is contiguous with the joint, attaches to the superior rim of the cricoid lamina between the two cricoarytenoid facets, and extends anteriorly to the medial surface of the arytenoid cartilage. Its primary function is most likely prevention of lateral dislocation of the arytenoids on forced abduction of the vocal folds.

The anterior ligament of the cricoarytenoid joint, extending from the vocal process of the arytenoids to the thyroid challenge, is the vocal ligament. When looked at from this standpoint, it can be viewed as an anterior check ligament axis for the arytenoid, capable of medial and lateral motion depending on the vector of forces applied to the arytenoid.

Motion permitted to the cricoarytenoid joint includes limited sliding along the long flat axis of the cricoid facet (only a few millimeters, as shown by Dickson and Maue-Dickson, 1982, on fresh cadavers). The primary motion is rocking (rotation) about the cricoarytenoid joint. This latter motion is quite free, with the ligaments acting as guide wires.

The vocal ligament itself is a condensation of the superior edge of the cricothyroid membrane (also known as the triangular membrane). The borders of the cricothyroid membrane can be defined as follows: superior border, vocal ligament; anterior border, anterior cricothyroid ligament; inferior border, superior rim of cricoid arch; posterior border, posterior end of vocal ligament, attaching to the vocal process and inferior fossa of the arytenoid cartilages. The body of this membrane is known as the conus elasticus and is the supporting structure for the deeper surface of the vocal fold below the ligament.

The vocalis muscle, the medial fibers of the thyroarytenoid muscle, travels posteriorly, with the vocal ligament forming the shelf-like body of the true vocal fold. As reviewed below, contraction causes bulking and shortening of the vocal fold and stiffening of the muscle and acts to help control pitch production.

ADDUCTION AND PHONATION CESSATION

In the normal larynx, the arytenoid cartilages can move over a relatively wide range of positions. Phonation can take place, however, within only a small portion of that range. The phonation portion has been called the "phonatory adductory range" and was found to be approximately only 14% of the full adductory range in a previous study (Scherer, 1995). It is expected that the range will be significantly restricted by abnormal conditions of the larynx.

Separating the arytenoid cartilages beyond the phonatory adduction range discontinues ongoing phonation. Another adductory method to stop phonation is to overadduct the arytenoid cartilages. Overcompression would be produced by high levels of contraction of the interarytenoid and lateral cricoarytenoid muscles, supplemented by membranous medial compression by the thyroarytenoid muscles. Overabduction would probably be governed by contraction of the posterior cricoarytenoid muscle. Intermediate levels of adduction and the resulting glottal configurations would depend on the antagonistic action of the adductory and abductory muscle systems (supplemented by the vocal fold lengthening muscle, the cricothyroid). Any obstructing pathologies in the glottis, such as tumors, papilloma, and membranous swellings, would alter effective adduction and glottal configuration potential.

PHONATORY THRESHOLD PRESSURE

Phonation requires a certain minimal amount of subglottal air pressure, the phonatory threshold pressure, to set the vocal folds into vibration. The minimal pressure to just maintain phonation is slightly less than the former. If the vocal folds are placed in the phonatory adductory range, the subglottal pressure must overcome the stiffness, mass, and internal tissue damping of the vocal folds in addition to any level of compression between the vocal folds to literally push them laterally and superiorly from below to start the first cycle. The lowest phonatory threshold pressure typically varies with fundamental frequency, ranging from ~3 cm H_2O (0.3 kPa) for lower pitches to ~6 cm H_2O (0.6 kPa) for higher pitches due to greater vocal fold tension (stretch) at higher pitches (Titze, 1994). There also may be an upper limit of subglottal pressure beyond which phonation is prevented or too chaotic to maintain normal voice quality.

Intervention with laryngeal surgery, voice therapy, and vocal training often attempts to regain normal phonatory threshold pressures or to lower existing thresholds, the latter creating less effortful phonation of the respiratory system. Establishment of lower phonation thresholds may correspond to more physiologically efficient phonation and should be an intervention goal.

PRESSURE EQUILIBRATION

If the translaryngeal air pressure (the difference between the subglottal pressure and the supraglottal pressure) were to become zero during speech, the pressure would be equal on all surfaces of the vocal folds, air would not pass, and phonation would not occur. This can be approached, for example, with overly prolonged voiced consonants such as */b/, /d/,* and */g/* during their corresponding (total) occlusion within the vocal tract (which may occur with certain speech pathologies and expressive or emphatic speech and singing). Prolonging the voicing of these consonants allows buildup of supralaryngeal air pressure until that pressure nearly equals the subglottal pressure, causing cessation of phonation as the translaryngeal pressure drops below the minimum sustaining pressure. Phonatory cessation in this manner may be part of a person's nonnormal speech production, highlighting an important articulatory-voicing interdependence.

GLOTTAL FLOW

Each phonatory cycle releases a time-varying glottal flow signal (also called the glottal volume velocity) that generates sound. **Fig. 44–1** shows two glottal volume velocity cycles of human phonation (top trace). The cycle period, from point A to D, is ~7.93 msec (126 Hz). The glottal volume velocity (usually given in liters per second, L/s, or cubic centimeters per second, cm^3/s) increases from point A to B during glottal opening, then decreases from point B to C during glottal closing. From point C to D the glottis is most closed. In this example there is a direct current (DC) flow offset of ~100 cm^3/s, produced by the presence of a posterior glottis opening throughout the cycle.

The lower trace of **Fig. 44–1** is the time derivative of the volume velocity signal of the upper trace. The fastest change of the volume velocity is at point E and corresponds to point E′ on the derivative waveform. Point E′ corresponds to the moment of time at which the greatest acoustic excitation is created.

In all voice intervention strategies, the attempt should be made to establish a suitable glottal flow waveform, because the voicing acoustics are related directly to the size, shape, and consistency (see later discussion) of the flow cycles. This is especially relevant in surgical cases in which the resulting structures of hard and soft tissue are atypical. The vibration of abnormal structures should yield normal flow waveforms if at all possible.

It is important to point out that the flow through the exit of the glottis at any moment in time is the accumulation of the air velocities over the cross section of the glottis. Indeed, the airflow may be faster through the anterior membranous portion of the glottis than the posterior membranous portion; that is, the velocity of the air is not the same across the glottal exit. Furthermore, the volume velocity waveform as seen in **Fig. 44–1** is obtainable from inverse filtering of the airflow that exits the mouth, which is a process that takes the airflow recorded at the mouth and removes the influences of the vocal tract resonances; that is, it is not a direct measure of the volume velocity at the glottis but is an inferred volume velocity waveform. Lastly, the glottal volume velocity, although highly useful, does not relate to all the sources of sound for vowels; it is highly likely that interaction between vortices and vocal tract boundaries

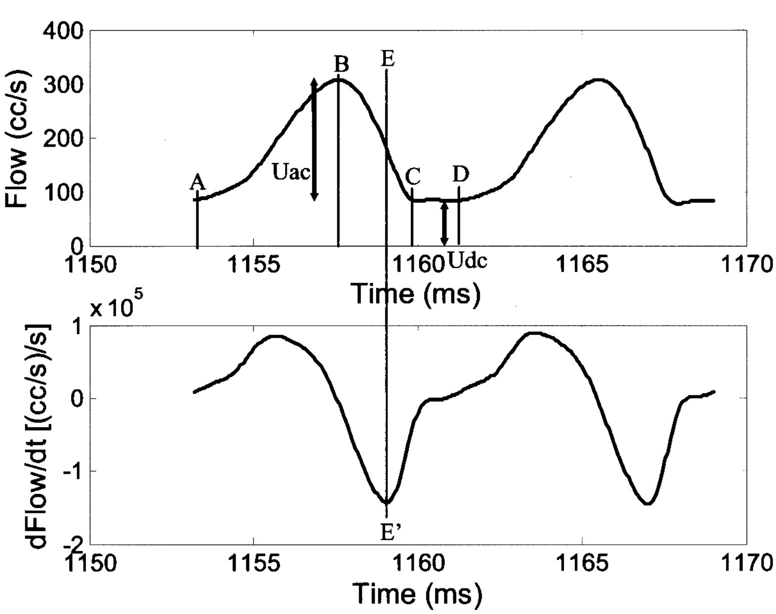


Figure 44–1 *Glottal waveform and its derivative. The top trace shows two cycles of glottal flow for a slightly breathy voice from a normal-sounding speaker. The phonation period for one cycle is the time between points A and D. The AC (alternating current) portion of the flow lasts from point A to point C. The DC (direct current) flow is the offset flow (Udc) and is seen from the zero line to the flow baseline. The magnitude of the AC flow (Uac) is from the flow baseline (seen from point C to D at ~100 cm^3/sec) to the peak flow at point B. The open quotient relative to the AC flow is the ratio of the time for the AC flow (point A to C) divided by the period. The lower trace shows the time derivative of the glottal flow. The moment of the maximum flow declination rate (MFDR) is shown along line E-E′. Skewing of the flow to the right can be seen by noticing that the time from point A to B is longer than the time from point B to C.*

create significant acoustic sources. Clearer understanding of the aeroacoustic interactions in voice and speech production is required and will likely enhance clinical intervention and voice training strategies. In this chapter, we will emphasize the relevance of the glottal volume velocity as shown in **Fig. 44–1**.

STRING MODEL FOR FUNDAMENTAL FREQUENCY OF PHONATION

Perceived pitch corresponds (nonlinearly) to the physical measure of fundamental frequency F_0. Fundamental frequency corresponds to the number of times per second the vocal folds complete a cycle of vibration. The string model for fundamental frequency, a simplified but useful model to conceptualize important factors in F_0 control, can be given by

$$F_0 = [1/(2L)](T/r)^{0.5},$$

where T is the tension of the vocal fold mass in motion, r is the density of the tissue, and L is the length of the vibrating vocal fold tissue. The model suggests that an increase of length or density, or a decrease in tension, of the tissue in vibration would result in a decrease in F_0. The string model equation is an excellent conceptualization in that it keeps clear the notion that tension, length, and density are all important concepts, and their relationship is essentially reflected by the expression.

PASSIVE F_0 CONTROL

When the vocal folds are stretched by contraction of the cricothyroid muscle, the tension within the different layers of the vocal fold tissue increases and the fundamental frequency tends to increase. This corresponds to the observation that the square root of the passive tension increases faster than the length increases in order for the F_0 to rise, as illustrated by the string model for F_0.

Pitch can be altered by conditions involving fluid engorgement, such as edema. Greater mass in vibration creates lower natural frequencies. This is seen through a manipulation of the equation above by recognizing that tension (T) is force (F) applied to the tissue divided by the cross-sectional area of the tissue in motion, and that density (r) is equal to the tissue mass divided by the corresponding tissue volume. Substituting these equivalences into the equation yields

$$F_0 = (1/2)\,(F/ML)^{0.5}$$

It can be seen, then, that doubling the mass due, say, to edema, all else the same, results in an F_0 value 29.3% lower; that is, a reduction of ~5 to 6 semitones. Indeed, this concept can be applied clinically to deduce approximately how much excess mass in the vocal folds a patient may have acquired.

Subglottal pressure plays a significant role in the control of pitch through passive tension increase of the vocal fold tissue in vibration. Data suggest that a change of 1 cm H_2O subglottal pressure results in an F_0 change of 3 to 6 Hz. As subglottal pressure increases, the maximum lateral excursion of the vocal folds increases, producing an increase in the stretch of the vocal folds within the cycle. This raises the average length and tension of the vocal folds during the vibratory cycle, increasing the F_0 value (passively) (Titze, 1994).

ACTIVE F_0 CONTROL

The above discussion emphasizes the contribution to pitch control through passive tension change of the vocal fold, whether through passive stretch by contraction of the cricothyroid muscle, dynamic stretch by increasing subglottal pressure, or mass changes as in vocal fold edema. The fundamental frequency is also dependent upon the activity of the vocalis (thyroarytenoid) muscle (Titze, 1994). The vocalis acts antagonistically to the cricothyroid muscle relative to length change of the cover (mucosa). If the cover only is vibrating, as in very soft phonation, increase in vocalis contraction should shorten and reduce the tension of the mucosal cover, thus lowering the fundamental frequency. If the vocalis muscle participates in the motion of the vocal fold to a significant degree, as in loud phonation, increase in vocalis muscle contraction will increase the effective tension of the entire tissue in motion, and thereby potentially raise the fundamental frequency (Titze, 1994).

At intermediate loudness levels, where the cover and vocalis muscle both participate in the vibration of the vocal folds, pitch may be controlled and even remain constant for many combinations of cricothyroid muscle contraction (to stretch the entire vocal fold passively) and vocalis muscle contraction (to shorten the cover and actively increase the tension of the vibratory vocalis portion). In addition, subglottal pressure can be used to control the pitch by varying the amount of vocal fold mass placed into vibratory motion and the amount of passive stretch tension applied to the vocal folds. It is clear that pitch control goes beyond the simple notion that contracting the cricothyroid muscle stretches the vocal folds and thus raises pitch. Rather, it is a complex, and not fully understood, combination of mass, length, active tensions, and passive tensions of the specific tissue in vibration, controlled by the glottal configuration, transglottal pressure, and muscle contraction levels.

Flexible use of subglottal pressure as a controller of both pitch and loudness (see later discussion) may help establish adequate stress patterns in speech and is therefore extremely important in the function of a limited larynx (e.g., paralysis or a conceptual pneumatic larynx implant). However, a fully expressive voicing source requires the complex interaction of factors already mentioned, and intervention strategies should attend to all factors.

PHONATORY INSTABILITIES

Perceptual judgments of an unclear or rough voice occur when the prominent moments of acoustic excitation during each cycle are not consistent from one cycle to the next. The time between primary acoustic excitations from one cycle to the next may vary (creating aperiodicities) if there are tissue abnormalities such as nodules, polyps, and unilateral stiffness abnormalities. This apparently causes kinematic (vocal fold motion) inconsistencies from cycle to cycle. Consecutively varying periods of primary acoustic excitations can be created by turbulent airflow through the glottis (as in breathy voice), creating added noise to the acoustic signal. Aperiodicities can be measured by jitter, one definition of which is the average cycle-to-cycle difference in period or equivalent fundamental frequency, with large jitter values corresponding to a greater sense of a lack of vocal clarity.

Pitch and vocal clarity are also affected by changes that occur over longer time lengths than a phonatory cycle. Diplophonia (the perception of two pitches simultaneously) and "subharmonics" (essentially integer subdivisions of the prominent fundamental frequency) come from multicycle length modulations of the volume velocity signal, giving rise to primary acoustic excitations at time intervals of twice (or more) the primary phonatory period. In vocal fry and "creaky" voice, pitch may be extremely low, dicrotic, or of varied roughness qualities, depending on the complexity of the low-frequency acoustic excitations. There are many kinds of measures for phonatory instabilities, and the reader is referred to the Suggested Readings for discussions and summaries.

SKEWING OF THE GLOTTAL FLOW

Fig. 44–2 is a schematic of a typical glottal volume velocity waveform with corresponding glottal motion. The glottal flow peak is produced after the peak of the

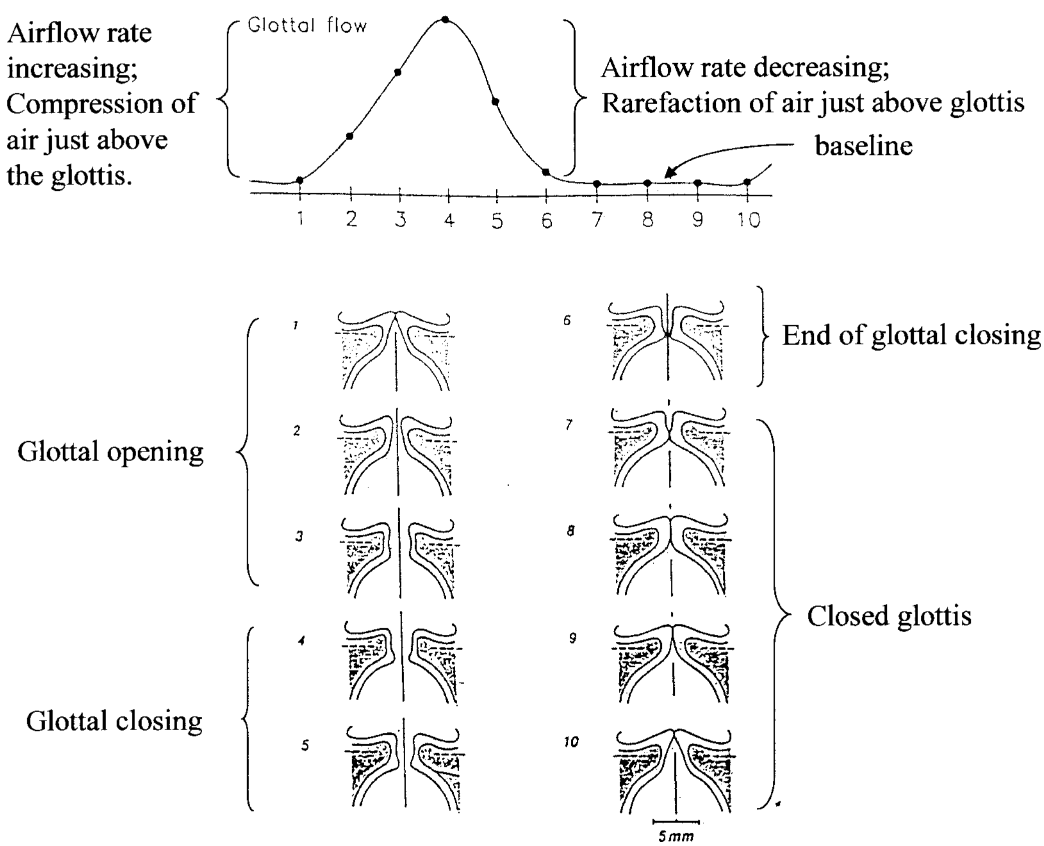


Figure 44–2 Schematic of the glottal flow waveform compared with the corresponding vocal fold and glottal motion.

glottal area occurs. This flow delay (or skewing to the right) characteristic is related at least to vocal tract inertance, amplitude of the glottal displacement, vertical glottal phasing, and potentially to the occurrence of the optimal glottal diffuser angle near the flow peak. The inertive effect refers to the fact that the air coming through the glottis must literally move other air already within the vocal tract, and this requirement slows the motion of the air as it first comes out of the glottis. After the moment of maximum flow, the airflow will reduce toward zero flow or to its minimum flow value as the two membranous vocal folds come together at the end of glottal closing (refer to **Fig. 44–1**). Because skewing of the glottal flow waveform results in a "brighter" voice, intervention strategies often need to be considered that might enhance the skewing of the glottal flow.

BREATHY, NORMAL, AND PRESSED PHONATION

The qualities of the voice depend on the size and complex shape of the glottal volume velocity waveform. Breathy voice is characterized by a more sinusoidal flow waveform than in normal phonation, with a significant flow bias. The flow bias (or "leakage") is unmodulated flow exiting the glottis due to the lack of complete glottal closure within the cycle. This leakage flow may create turbulence and add inharmonic spectral components to the flow waveform, giving the sound the "breathy" quality. In contrast, in pressed phonation, the peaks of the flow are significantly smaller than for normal phonation, and the amount of time within the cycle during which the air exits the glottis is relatively short compared with normal phonation at the same pitch. Spectrally, breathy quality has energy primarily in the first few partials, along with aperiodic noise components, whereas greater energy is distributed to higher partials in normal voice quality. In pressed voice, the energy in the first two partials is reduced.

SPECTRAL ASPECTS OF THE GLOTTAL FLOW

The intensity and spectrum of vocal sound depend on the glottal volume velocity waveform. The overall intensity or sound pressure level (SPL) of the output sound will increase as the maximum rate of change of the glottal volume velocity "shut-off" (line E-E′ in **Fig. 44–1**) increases. A greater maximum slope of the glottal airflow waveform has the spectral effect of raising the energy of the partials primarily within the region of the first formant, usually the highest spectral portion for overall SPL.

It is of importance to note some other spectral effects of shaping differences of the glottal volume velocity waveform. The greater the AC flow amplitude of the volume velocity waveform, the greater the amplitude of the fundamental frequency. A doubling of the amplitude of the waveform corresponds to an increase of ~3 to 7 dB in the spectral level of the fundamental frequency. When the flow has nearly reached its lowest value near baseline, a "shut-off corner" is created. A very sharp corner compared with a well-rounded corner can cause the intensity of the upper partials to increase by up to 10 to 20 dB, undoubtedly affecting the quality of sound. For example, changing glottal adduction from a normal voice quality to a breathy quality would round off this corner considerably. This concept needs exploration relative to how the vocal folds come together and to the perception of vocal quality.

An important variable controlling intensity and spectra of the glottal airflow is the subglottal pressure. As subglottal pressure increases for a constant level of glottal adduction, the peak airflow through the glottis increases. As the maximum value of the volume velocity waveform increases, the greater is the intensity level of the fundamental frequency, as already discussed; also, the maximum flow derivative should increase if the time during which the flow decreases remains the same. In addition, the increase in subglottal pressure may cause the vocal folds to come back together faster (or perhaps alter their dynamic phasing) after their maximum excursion, creating a sharper flow shut-off corner near baseline, raising the spectral level of the higher partials, as discussed earlier. Therefore, greater subglottal pressure may increase the flow peak, increase the maximum flow derivative, and sharpen the baseline flow shut-off corner. These effects change the flow spectrum shape by increasing the intensity of the voice. A doubling of the difference between subglottal pressure and threshold pressure should raise the source acoustic power by ~6 dB.

Intensity is affected strongly by the fundamental frequency of voice production. The glottal power output should increase by 6 dB for an octave rise in fundamental frequency (all else remaining the same), due to the dependence of the maximum flow derivative on the fundamental frequency. The higher frequency for women compared with men (a ratio of ~1.7:1) is offset by a larger amplitude of glottal volume velocity for men (a ratio of ~2:1), so that the SPL for women is only 1 to 2 dB lower than for men (continued studies of these comparisons are necessary).

INTRAGLOTTAL PRESSURES AND VOCAL FOLD OSCILLATION

The mechanical phonatory motion of the vocal folds depends on the folds being driven by air pressures within the glottis. The dynamic (time varying) air pressure forces work with the biomechanical characteristics of the vocal folds (mass, stiffness, damping) to overcome the damping losses within the tissue to permit vocal fold vibration. If the intraglottal pressures are negative, they exert a pulling force to bring the vocal folds together. If the intraglottal pressures are positive, they exert a force to separate the two vocal folds.

The glottis takes on two primary shapes during phonation, convergent and divergent. During glottal opening, the convergent glottal shape is produced with a wider separation at glottal entrance and a narrower separation at glottal exit (**Fig. 44–2**). The glottal convergent shape, together with the positive pressure in the subglottis, creates a positive pressure within the glottis, and this positive pressure pushes on the vocal folds, facilitating glottal opening. During glottal closing, the divergent glottal shape may be prominent, with a narrower opening at glottal entrance than at glottal exit. Due to this divergent shaping, the pressure at glottal entry is lower than at exit and is therefore negative within the glottis if the glottal exit pressure is atmospheric (zero) or negative. This negative pressure due to the divergent shape of the glottis is consistent with the Bernoulli energy equation and pulls on the vocal fold surfaces, facilitating glottal closure.

Fig. 44–3 illustrates the intraglottal pressure variations for different convergent and divergent glottal angles. These pressure distributions were obtained using a Plexiglas model with numerous glottal pressure taps. The convergent glottis creates positive surface pressures on the vocal fold medial walls. The uniform glottis (where the glottis diameter is constant from entrance to exit) has positive pressures also. Divergent shapes, however, create negative pressures within the glottis, especially near the entrance. The greatest negative pressures occur when the intraglottal divergent angle is between 5 and 10 degrees. This corresponds to near "optimal" diffuser angles, where pressures are most negative and glottal flow resistance is lowest (thus potentially contributing to the peak glottal flow occurring after the peak glottal area has occurred). A primary goal of surgical, pharmacological, and behavioral therapies, if feasible, should be to restore the ability of the vocal folds to take on the convergent-to-divergent glottal shapes necessary to allow the intraglottal pressures to alternate between positive and negative pressures. If this alteration is not possible, it is doubtful that phonation can occur or be maintained if initiated.

LARYNGEAL FLOW RESISTANCE

For a specific adduction level (vocal process gap), the relation between average translaryngeal pressure and average glottal airflow has been found to be linear when using excised canine larynx models (Alipour et al, 1997), a finding consistent with an unpublished human phonation study. As adduction increases [i.e., as the vocal process gap (VPG) decreases], the linear relations shift. **Fig. 44–4** schematizes these findings. As the level of adduction increases, the pressure-flow line moves to the left in

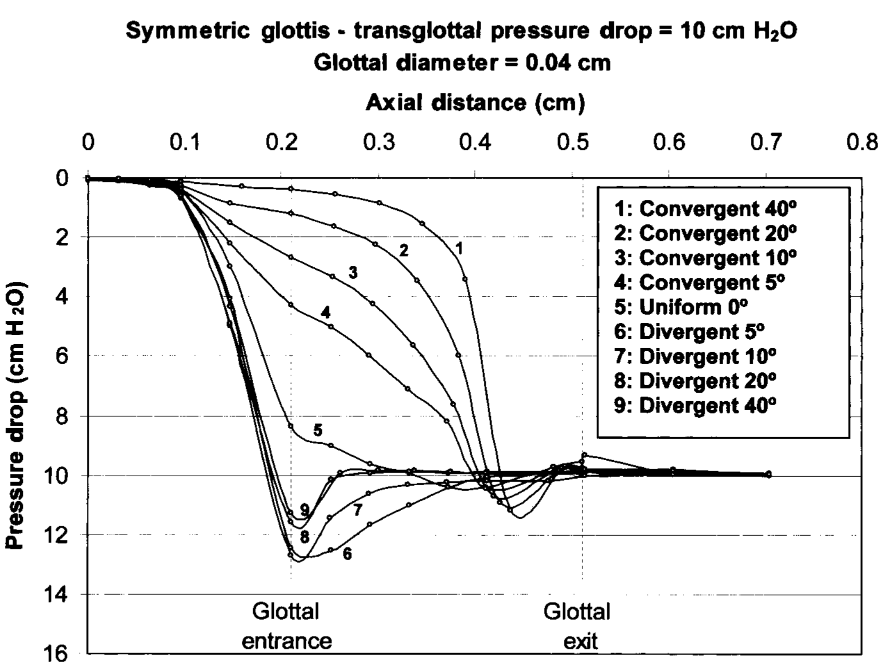


Figure 44–3 Intraglottal wall pressure distributions. Each line is for a specific glottal angle. All cases are for a transglottal pressure drop of 10 cm H_2O and a minimal glottal diameter of 0.04 cm. The points on the lines are empirical data from the use of a Plexiglas model of the larynx. If one considers the pressure at glottal exit to be zero, then convergent glottal angles yield positive pressures in the glottis up to the rounded glottal exit location, and divergent glottal angles yield negative pressures in the first half or so of the glottis.

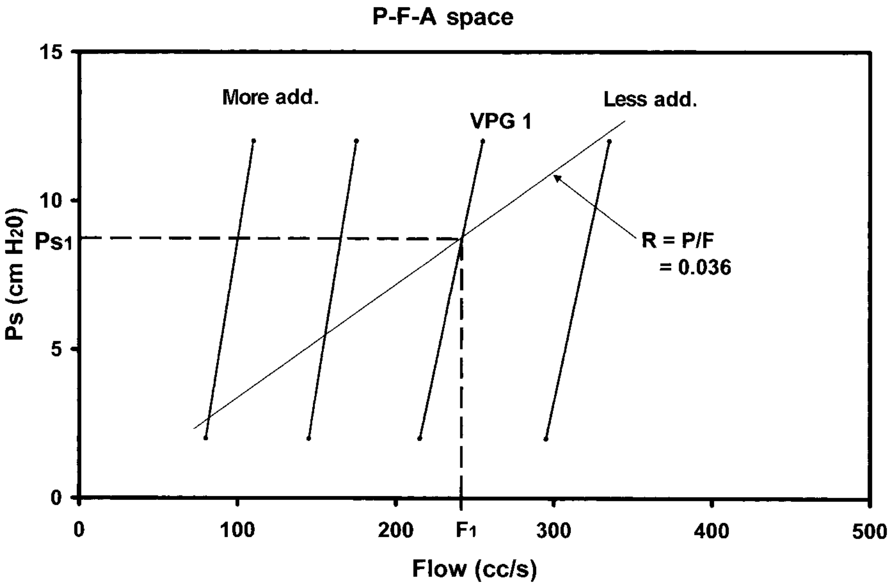


Figure 44–4 *Pressure-flow-adduction (P-F-A) space. The schematic illustrates the finding that any particular value of laryngeal flow resistance (translaryngeal pressure divided by glottal volume flow) can be obtained for a wide range of glottal adduction values. Also, if this space were obtained for a patient, it may be possible, for a particular phonation, to predict the level of glottal adduction by knowing both the translaryngeal pressure and the flow for that phonation.*

the figure. Laryngeal flow resistance is usually defined as the ratio of the average translaryngeal pressure to the average laryngeal airflow. One reason that laryngeal flow resistance is often used in studies of phonatory behavior is the desire to measure glottal competency, wherein the assumption is that, if flow resistance is low or high, glottal adduction values will be low or high, respectively. **Fig. 44–4** illustrates a caution to this generalization. The broad diagonal line represents a constant value of flow resistance, suggesting that a significant range of adduction values (not a narrow range of adduction) can yield the same value of flow resistance. Reports of laryngeal flow resistance must be accompanied by the report of pressure or flow to allow for a more appropriate interpretation of the resistance values relative to adduction.

VOCAL PHYSIOLOGY IN THE ABNORMAL STATE

Benign mucosal pathology of the true vocal folds is brought about through a myriad of etiologic factors. These may include biomechanical issues (e.g., habitual hard glottal onset or poor abdominal support with concomitant excessive use of the perilaryngeal musculature), irritation (e.g., cigarette smoke, esophageal reflux, infected postnasal drip, or exposure to allergens), voice abuse (e.g., shouting while ill with acute viral laryngitis), and endocrinologic or other systemic illness (e.g., hypothyroidism or amyloidosis).

The degree of perceptual symptoms produced depends more on the stiffness or mass of the vocal folds, the degree of contact, and the amount of breathiness produced, rather than on the actual type of mucosal pathology. Of importance from the standpoint of mucosal disease is the layer(s) of the vocal fold that are involved.

Hirano (1981) has described the true vocal fold on coronal section in terms of cover, transition, and body. By definition, the cover consists of the overlying stratified squamous epithelium and the superficial layer of the lamina propria (also known as Reinke's space). Considerable recent interest has been given to the basement membrane zone, which provides physical support to and helps repair the epithelium while anchoring it to Reinke's space. This layer consists of ground substance including glycosaminoglycogen and glycoproteins, collagen and elastin, allowing for a loose, pliable consistency.

Anatomically, Reinke's space is sharply delimited by dense, fibrous tissue at the anterior commissure, along the vocal process of the arytenoids, and beneath the free margin of the true vocal fold. Its upper limit is not as well defined and may vary considerably, usually reaching to the inferior aspect of the ventricle. This anatomical relationship has significant implications for the disease process of Reinke's edema, which is discussed later.

The intermediate and deep layers of the lamina propria consist predominantly of elastin and collagen, respectively. They form the vocal ligament and are called the transition by Hirano. The deep layer of the lamina propria merges into the underlying vocalis muscle (called the body by Hirano).

Most benign processes only involve the cover (stratified squamous epithelium, basement membrane zone, and Reinke's space). Many benign processes, including mature nodules, cysts, scar, sulcus vocalis, and some polyps, will increase the stiffness of the cover. Reinke's edema and some polyps may actually decrease the stiffness of the cover, while increasing the mass however.

Carcinoma, once it invades the transition, will increase stiffness not only in the cover but also in the transition. The stroboscope is extremely sensitive in detecting changes caused by fixation or involvement of the ligament from small carcinomas. The authors use the stroboscope routinely to examine patients with potential carcinoma. Rigidity caused by dysplasia or early carcinoma usually can be identified before the tumor has caused vocal fold immobility.

Mass of the cover is increased by certain benign groups of pathologies (e.g., nodules, polyps, Reinke's edema). It may be decreased in sulcus vocalis or scar.

Reviewing certain benign vocal fold pathologies, vocal fold polyps have a wide variation of histological appearance, ranging from predominantly gelatinous, with a loose edematous stroma and spare collagen fibers, to predominantly telangiectatic, with labyrinthine, sinus-like channels. The base of a polyp can, at times, involve the ligament.

Polyp formation has been linked to such etiologies as phonotrauma with mechanical stress causing localized subepithelial edema, development and abrupt release of high subglottic pressure, and increased hyperemia of a vocal fold hemorrhage. The etiology of some is unknown.

Polyps generally are unilateral, but they can vary greatly in size and shape and can be pedunculated or sessile. Acoustic analysis of voice production from individuals with polyps has demonstrated that the size of the polyp is correlated with several acoustic measures, including fundamental frequency (a lowering thereof), roughness of the voice, asymmetry and irregularity of vocal fold vibration, and the pitch and amplitude perturbation quotient. The degree of glottal gap correlates negatively with maximum phonation time and sound pressure level, but positively with mean airflow rate, as would be intuitively expected. Stiffness of the cover varies in polyps. It increases when the main feature is hyaline degeneration but decreases when it is edema. The increased mass of the cover associated with polyps causes disturbances of periodicity and vibratory synchrony.

Vocal fold nodules generally are caused by voice abuse. They begin as edema and vasodilatation (prenodular diathesis). In this instance there is increased mass of the cover but little change in stiffness. The mucosal wave is intact in the chest register, but on stretching of the vocal folds in head register, there is "hourglassing," with both a posterior and an anterior gap seen on stroboscopy. Clinically, this corresponds to a breathy quality of the voice with loss or instability of the high frequencies, and frequently with instability of the *passaggio*. In well-established nodules, the histology demonstrates dense collagen deposition, similar to epithelial callus. In these circumstances, both the mass and the stiffness of the cover are increased and the "hourglass" shape of the glottis is more obvious. Clinical findings are similar to but more pronounced than those in the prenodular diathesis scenario.

A caveat in singers, as per Sataloff (1997), deserves mentioning. Many singers develop bilateral symmetrical soft swellings at the junction of the anterior and middle thirds of their vocal folds following heavy voice use. There is no evidence to suggest that singers with these "physiological swellings" are predisposed to the development of vocal nodules.

Reinke's edema is an accumulation of mucoid fluid in the superficial layer of the lamina propria (Reinke's space). This is caused by repeated irritation or injury to the larynx, frequently in the clinical backdrop of heavy cigarette use and reflux esophagitis. Hypothyroidism may play a role as well. The edema tends to be bilateral and diffusely located throughout the space. The outstanding histological finding is lakes of mucoid fluid interwoven with sheets or masses of immature elastic fibers. There is a marked increase in the mass of the cover but decreased stiffness. The ligament is not involved.

Clinically, Reinke's edema causes a low, gruff, husky voice that correlates acoustically with an abnormally low mean speaking fundamental frequency and severely reduced dynamic range. In the more severe cases, the highest sustainable tone is at least one octave below normal. In the most severe cases, there is marked increase in the size of the vocal fold, occasionally to the extent of ball valving of tissue upon respiration.

Sulcus vocalis refers to a fine furrow or invagination of the covering epithelium along the length of the membranous vocal fold. Vergeture describes an atrophic depression along the free margin, the medial margin of which is bound. Not infrequently, these two descriptors are clinically interchanged, although vergeture is, in fact, not a true sulcus. Incomplete closure occurs during phonation in both conditions. The sulcus represents a deficiency of tissue in the cover. Histologically, it is a blind sac that extends down to and adheres to the vocal ligament. Mechanically, the mass of the cover is reduced, and stiffness is markedly increased. Acoustically, the voice is breathy, husky, and loud, and the phonation time is short.

Because sulcus vocalis is long-standing, secondary phonatory behavioral patterns are developed by effectively all patients with this spectrum of disorders, including false vocal fold phonation. This leads to an array of acoustical changes caused by the associated hyperfunction.

This section will not review the acoustical abnormalities subsequent to the various hyperfunctional voice disorders (Morrison's muscular tension dysphonias). Muscular tension dysphonias are commonplace, however, and represent the majority of pathologies seen in our voice clinic. In fact, mucosal pathology of the vocal folds often develops secondary to a musculoskeletal problem. The interested reader is referred to Morrison et al (1994) and to Harris et al (1998).

SUMMARY

This chapter has covered the basics of voice production, focusing on phonatory behavior. It has purposely given little attention to other critical functions of the larynx, such as airway protection and pressure valving. Rather, it has reviewed phonatory behavior, predominantly in the normal state. Biomechanics of the larynx has been emphasized, as have such aspects as phonation cessation, phonatory threshold pressure, pressure equilibrium, glottal flow and skewing thereof, fundamental frequency control, phonatory instabilities, breathy, normal, and pressed phonation, spectral aspects, intraglottal pressures and laryngeal flow resistance, and vocal fold oscillation.

The chapter has reviewed certain aspects of vocal physiology in the abnormal laryngeal state, including such benign mucosal pathologies as vocal polyps, nodules, Reinke's edema, and sulcus vocalis. The effects of changes in mass and stiffness of the vocal folds have been correlated with general acoustic and clinical findings.

ACKNOWLEDGMENTS

Portions of this chapter were adapted from Scherer RC, Rubin JS. Laryngeal physiology: normal and disordered. In: Benninger M, ed. Benign Disorders of the Voice. Alexandria, VA: American Academy of Otolaryngology—Head and Neck Surgery Foundation; 2002:29–44.

This chapter was supported in part by grant R01 DC03577 from the National Institute on Deafness and Other Communication Disorders.

Suggested Readings

Alipour F, Scherer RC, Finnegan E. Pressure-flow relationships during phonation as a function of adduction. J Voice 1997;11(2):187–194

Baer T, Sasaki C, Harris KS. Laryngeal Function in Phonation and Respiration. San Diego: College-Hill Press; 1987

Baken RJ. Clinical Measurement of Speech and Voice. Boston: College-Hill Press/Little, Brown; 1987

Bless DM, Abbs JH. Vocal Fold Physiology, Contemporary Research and Clinical Issues. Boston: College-Hill Press; 1983

Dickson DR, Maue-Dickson W. Anatomical and Physiological Bases of Speech. Boston: Little, Brown; 1982

Fujimura O, ed. Vocal Physiology: Voice Production, Mechanisms and Functions. New York: Raven Press; 1988

Gauffin J, Hammarberg B. Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms. San Diego: Singular Publishing Group; 1991

Harris T, Harris S, Rubin JS, Howard DM, eds. The Voice Clinic Handbook. London: Whurr Publishers; 1998

Hirano M. Clinical Examination of Voice. New York: Springer-Verlag; 1981

Kirchner JA, ed. Vocal Fold Histopathology: A Symposium. San Diego: College-Hill Press; 1986

Laitman JT, Reidenberg JS. Specializations of the human upper respiratory and upper digestive systems as seen through comparative and developmental anatomy. Dysphagia 1993;8(4):318–325

Morrison M, Rammage L, Nichol H, et al. The Management of Voice Disorders. San Diego: Singular Publishing Group; 1994

Rubin JS, Sataloff RT, Korovin GS, Gould WJ, eds. Diagnosis and Treatment of Voice Disorders. New York: Igaku-Shoin Medical Publishers; 1995

Sataloff RT, ed. Professional Voice: The Science and Art of Clinical Care. 2nd ed. San Diego: Singular Publishers; 1997

Scherer, RC. Laryngeal function during phonation. In: Rubin JS, Sataloff RT, Korovin GS, Gould WJ, eds. Diagnosis and Treatment of Voice Disorders. New York: Igaku-Shoin Medical Publishers; 1995:86–104

Stevens KN, Hirano M. Vocal Fold Physiology. Tokyo: University of Tokyo Press; 1981

Sundberg J. The Science of Singing. Dekalb: Northern Illinois University Press;1987

Sundberg J, Titze I, Scherer R. Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. J Voice 1993;7(1):15–29

Titze IR. Vocal Fold Physiology: Frontiers in Basic Science. San Diego: Singular Publishing Group; 1993

Titze IR. Principles of Voice Production. Englewood Cliffs, NJ: Prentice-Hall; 1994

Titze IR, Scherer RC. Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control. Denver: Denver Center for the Performing Arts; 1985

Wyke B. Ventilatory and Phonatory Control Systems: An International Symposium. Oxford: Oxford University Press; 1974

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. Critical functions of the larynx include which of the following?
   A. Protection of the lower airway from food particles
   B. Voicing
   C. Sudden increase in intra abdominal and intrathoracic pressure
   D. A and C only
   E. All of the above
2. All of the following are true regarding the biomechanics of the larynx, except
   A. Each of the cricoarytenoid joints has an elliptical facet measuring ~9 mm (in adult males).

B. Dickson and Maue-Dickson identified a stretch of 25% in the length of the vocal ligament in fresh cadavers via contraction of the vertical belly of the cricothyroid muscle.
C. The primary function of the posterior cricoarytenoid ligament is the prevention of lateral dislocation of the arytenoids on forced abduction.
D. The vocal ligament is a condensation of the superior edge of the cricothyroid membrane.

3. The minimal absolute threshold of subglottic air pressure typically required to set the vocal folds in motion varies with fundamental frequency. It ranges from about _____ cm H_2O for lower pitches to about _____ cm H_2O for higher pitches (fill in the blanks).
   A. 3, 6
   B. 6, 9
   C. 9, 12
   D. 12, 16
   E. 15, 20
4. Contraction of which muscle, in soft phonation, is likely to act antagonistically to the cricothyroid muscle and thereby lower pitch?
   A. Lateral cricoarytenoid
   B. Interarytenoid
   C. Thyroarytenoid
   D. Posterior cricoarytenoid
   E. Thyroepiglottic
5. In pressed phonation, which of the following is/are true?
   A. The peaks of the glottal flow waveform are significantly greater than for normal phonation.
   B. The amount of time within the cycle during which air exits the glottis is relatively long compared with normal phonation (at the same pitch).
   C. Spectrally, the energy in the first two partials is reduced.
   D. A and C only
   E. All of the above
6. A doubling of the difference between subglottal pressure and threshold pressure should have what effect on the source acoustic power?
   A. It should lower it by 6 dB.
   B. It should lower it by 3 dB.
   C. It should have no effect.
   D. It should raise it by 3 dB.
   E. It should raise it by 6 dB.
7. Which of the following is true about the pathological condition sulcus vocalis?
   A. Histologically, there is an accumulation of mucoid fluid in the superficial layer of the lamina propria.
   B. Histologically, it is defined by telangiectatic stroma with labyrinthine sinus-like channels.
   C. On stroboscopy an “hourglass” appearance is seen, with both an anterior and a posterior gap.
   D. Mechanically, the mass of cover is reduced, but stiffness is markedly increased.
   E. It is associated clinically with a low, gruff, husky voice correlating with an abnormally low mean speaking fundamental frequency.

Chapter 45

PRINCIPLES OF PHONOSURGERY

Peak Woo

Phonosurgery is surgery performed primarily for improvement of the voice. The term *phonosurgery* was used by Von Leden in the 1950s to describe any surgery performed specifically to improve voice. Today it encompasses a wide variety of procedures. Some of these include phonomicrosurgery of the vocal folds, augmentation laryngoplasty for vocal fold paralysis, and pitch change by laryngeal framework surgery. A partial list of the types of phonosurgery procedures performed today is given in **Table 45–1**. Undoubtedly, the list will change as new surgical procedures are invented to address voice problems. Although phonosurgery procedures may change, the principles of phonosurgery are based on anatomy, physiology, pathology, and pathophysiology. There are some specific aspects of anatomy and physiology that are relevant to the performance of phonosurgery. The resident and practitioner interested in voice and phonosurgery should be versed in the basics of acoustics, vocal fold anatomy, and physiology. The purpose of this chapter is to provide the basic science principles guiding phonosurgery. Besides aiding the surgeon in understanding today's phonosurgery techniques, basic knowledge about the anatomy, physiology, and pathophysiology relevant to voice production will be helpful in understanding any future developments in phonosurgery. With this knowledge, the surgeon should be better able to diagnose the extent of functional deficit and perform the phonosurgery to maximize function.

THE MAIN ASPECTS OF PHONATION

The method of producing the tone for sound resides in the glottis. Air passes through the vocal folds, which are alternately closed and opened by membranous tissue. This stream of air from the chest gets separated into individual pulses that vibrate the vocal folds. The

TABLE 45–1 PHONOSURGERY PROCEDURES

Type of Surgery	Procedure
Laryngoplastic phonosurgery	Medialization laryngoplasty Arytenoid adduction Lateralization laryngoplasty Cricothyroid approximation Mucosal grafting
Phonomicrosurgery	Laser microlaryngoscopy Microlaryngoscopy and excision Cordotomy and microflap surgery
Injection laryngoplasty	Fat injection Teflon injection Gelfoam injection Steroid injection Botox injectioin
Neuromotor surgery	Recurrent nerve section Muscle myectomy

vocal folds are a special set of vibrating bodies because they are made of elastic membranous tissue that is subject to deformation by the airstream. Two examples of vibrating tissues that are deformed by the airstream are the lips during the playing of brass instruments and the vocal folds or during singing. Unlike vibrating bodies made of stiff, concompliant materials, such as reeds and strings, the pressure required to bring vibrating bodies such as vocal folds into oscillation is low. The vocal folds need very little lung pressure to overcome the resistance offered by the vocal folds. Thus, in the human larynx, vocal fold oscillation may be sustained at pressures as low as 4 cm of water. This makes phonation an easy, low-pressure system in the normal state. When this low phonation state is disturbed by pathology, one of the major disturbances is a higher pressure requirement of speaker to produce sound. This makes the effort to speak harder.

In the human larynx, evolution has made remarkable changes in the proportions of the membranous larynx to cartilaginous larynx. Thus the adult membranous larynx is predominant and occupies the anterior two thirds of the glottal proportion. This has two advantages for the production of sound. One, the vocal folds act as large membranous tongues that can be altered in configuration, tension, and form by the airstream. Large membranous folds can be altered by adjustment of the vocal fold aperture by muscular voluntary control. A longer membranous fold allows for better differentiation of insertion of the muscles into the vocal ligament, allowing greater rapidity and extraordinary certainty. Secondly, the larynx has progressively lowered into a more caudal position. This gives the resonating vocal tract greater length. A resonating tract, which is adjustable and long, has advantages of short, rigid tubes. This makes the vocal tract an ideal instrument for adjustment. Adjustments of the vocal tract allow the vocal tract to amplify certain frequencies selectively and make sounds with greater tonal richness in both speaking and singing. Thus we see evolutionary changes to the vocal tract that are optimized for vocal communication.

PHYSIOLOGY OF PHONATION

ACOUSTICS AND THE SPEECH PRODUCTION CHAIN

Communication is heavily reliant on ideas expressed through speech. Its effortless production in normal states is often taken for granted in phone and personal conversations. The ideas and tonal qualities during vocal expression give voice a window into a person's character and personality. With the increasing importance of computers and technology, machines will undoubtedly be managed by voice and speech activation. The need to communicate spans the life cycle. With the increasing aging of many populations, including in the United States, speech and voice communications will often suffer as part of senescence, creating a major health problem in communication and quality of life in later life.

Voice and speech production is the end result of linguistic formulations by the brain and the motor sequence, which are sent to the muscles of the diaphragm, the larynx, the tongue, and the lips. The neural impulses coordinate the right amount of diaphragmatic pressure to push the air through the larynx, while the vocal folds serve as a flow converter. The vocal folds oscillate, changing airflow from the steady direct current (DC) airflow into the pharynx. The rapid oscillation of the vocal folds cuts the steady airstream into many small individual puffs of air in the pharynx. The larynx has been compared to a DC to alternating current (AC) flow converter. The by-product of airflow conversion is rapid compression and rarefaction of the airstream, thus creating minute changes in pressure in the air. These sound pressure variations are called a sound wave. As the vocal folds oscillate, the sound waves are generated at the glottis. The sound wave is propagated from the speaker's larynx to the ear of the listener. The sound wave propagated through the human vocal tract is subject to further amplification and damping by the supraglottic larynx, pharynx, oral cavity, and lips. The vocal tract modifies the glottal sound to produce the final sound as it exits the speaker's vocal tract. Thus the production of sound may be seen as a series of mechanical, aerodynamic, aeroacoustic interactions that produce sound by the fluttering of the vocal folds and amplify sound by the supraglottic subsystems (**Fig. 45–1**). The vocal mechanism is conveniently divided into the three subsystems: source (lungs, diaphragm, and abdominal control), converter (larynx and vocal folds), and filter and amplifier (resonator tract of the pharynx, oral cavity, lips, etc.). The diagnostic challenge to the otolaryngologist faced with a patient with dysphonia and speech disorder is to differentiate first the various subsystems affected by pathology or pathophysiology. After the affected subsystem is identified, further investigation is performed to evaluate the pathology using instrumentation and physical examination. Often, one or multiple involvements by disease or function may affect any or all three subsystems. This will be discussed separately.

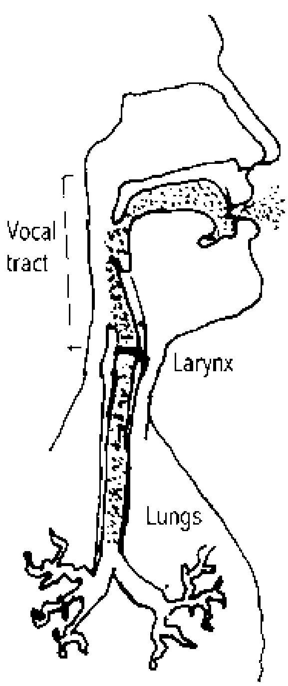


Figure 45–1 *Vocal tract is divided into subsystems of pulmonary power source, sound source, and resonator. An analogy is the lungs and respiratory system are the power generator, the glottis source is the direct current (DC) to alternating current (AC) flow converter, and the vocal tract above the glottis serves as the magnifier.*

Respiration, Phonation, and Phonatory Airflow

The source of energy for voice production is generated by the pulmonary system. The muscles necessary to propel airflow consist of the muscles of the thorax and diaphragm, and the extrathoracic muscles anchoring the thorax to the skeleton. For adequate support of the voice, both pulmonary pressure and an adequate pulmonary air reserve are critical. Poor coordination of the respiration and phonation systems may occur in neurological systems such as head trauma, stroke, and muscle diseases (myasthenia gravis). Traditional volume–time spirograms can be used to measure the forced vital capacity (FVC), forced expiratory volume (FEV), and peak expiratory flow (PEF). These spirograms are good measures of lung function and should be utilized as part of a phonatory function test. This is especially true when there is a suspicion of lung dysfunction. Phonatory airflow is, however, not directly measured by pulmonary function tests (PFTs) or by spirometry tests. For measures of phonatory airflow, several methods are available but not commonly used. A handheld spirometer can be used to calculate the mean phonatory volume (MPV) per breath. MPV is less than vital capacity and is usually 2000 to 3000 cc. This is often effort, age, and gender dependent. Reduction in these functions may be present in the patient with chronic pulmonary disease, restrictive lung diseases, and other chronic illness. With the addition of a stopwatch, the phonatory volume divided by the phonation time gives the mean flow rate (MFR). The MFR is a mean value of the airflow rate traversing the vocal folds during sustained vowel phonation. This value may also be measured directly or indirectly by other means such as pneumotachography, hot wire anemometry, and inverse filtering of the flow signal. The MFR for males is slightly different and varies somewhat by age and phonatory token used for testing. MFRs done at our institution are 100 to 180 cc/sec in males and 80 to 150 cc/sec in females for the vowel *ee*. The variation in mean flow in normal subjects shows that MFR is a relatively insensitive test of phonatory function. It is useful in grossly disturbed glottal closure situations such as vocal cord paralysis and bilateral vocal cord paralysis and in the study of neurogenic dysphonia with glottal incompetence.

The evaluation of respiratory dynamics is especially valuable in vocal pedagogy and studies of the extraordinary voice. Deep respiration and a coordinated respiration to phonation pattern are critical to deep relaxation and a production of singing that is free from strain and muscle tension. In patients with normal larynges and lungs, poor use of the pulmonary system can result in functional disorders resulting in dysphonia. A common type of functional dysphonia seen by otolaryngologists is muscle tension dysphonia. Many patients with functional dysphonia will have poor respiratory dynamics and phonation coordination. Understanding respiration and breath support of voice is one of the keystones in the rehabilitation of the patient with functional dysphonia.

Fold Vibration

Phonation is the creation of the glottal sound source by vibration of the glottal source. The vocal folds acts as a DC to AC flow converter. A change of the steady breath stream from the lungs into a highly oscillatory, alternating flow signal is one of the main mechanisms of the vocal folds to generate the glottal sound (**Fig. 45–2**). Normal vocal folds are capable of great variations in the production of different vibratory patterns. One simplistic account of the process of vocal fold vibration is the following. First, the buildup of the subglottic pressure generated by the lungs is increased to phonation threshold pressure by the adducted vocal folds; second, the subglottic pressure pushes the vocal folds apart and deforms the soft membranous vocal folds; third, there is a sudden release of airflow as the vocal folds are pushed aside, creating the jet of airflow traversing the glottis at high velocity; fourth, this release of air releases the subglottic pressure, allowing the elastic recoil of the vocal folds to reapproximate and close the airway; fifth, the approximated vocal folds again build up the subglottic pressure to allow the cycle to start again. This cycle is repeated at hundreds of times per second in a quasiperiodic manner. The subglottic pressure, the degree of approximation of the vocal folds, and the airflow rate dictate the amplitude and the frequency. The frequency of vibration of the vocal folds is directly related to the fundamental frequency of the sound spectrum. The amplitude of vocal fold oscillation is related to the loudness and affects the acoustic spectrum of the glottal sound.

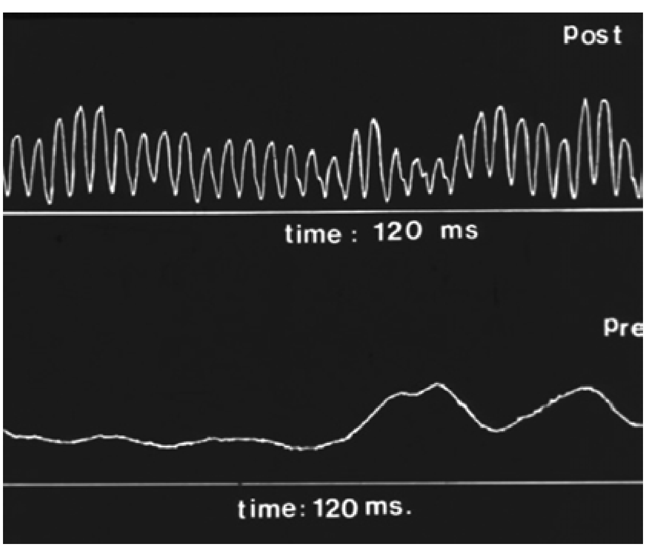


Figure 45–2 *The concept of DC to AC flow converter is shown in this illustration. The airflow is sampled at the mouth at high frequency. This was done before and after surgery for papilloma. Before surgery, DC flow with little modulation. Postsurgery, good airflow with high AC component. Note that the mean phonatory flow rate is the same.*

There are many forces active in the oscillation of the vocal folds. The myoelastic aerodynamic theory of phonation by Van den Berg states that phonation is a process that depends on the interaction of an elastic valve, air pressure, and airflow. Some of the physical parameters important to vibration include the elastic nature of the vocal folds, the subglottic pressure driving the opening of the folds, the viscosity of mucous and elastic forces achieved by tissue deformation, and the Bernoulli effect. Thus the vocal oscillator has been described in terms of force, resistance, pressure, displacement, compression, velocity, and other physical measures.

The determinant variables, which dictate vibration at set frequency and amplitude, include the driving pressure (subglottic pressure) and the viscoelastic properties of the vocal folds. These two forces are operating in an antagonistic manner. The subglottic pressure will push the vocal folds apart, whereas the viscoelastic, aerodynamic forces of flow will bring the vocal folds together. The viscoelastic properties of the vocal folds may be further broken into dependent variables such as length of the vibratory body, mass of the vocal fold, and compliance of the oscillating body. Thus length, mass, and tension forces act on each vocal fold during vocal fold oscillation to determine the frequency, amplitude, and periodicity of vocal fold oscillation. Although rarely are they acting independently in the disordered larynx, it is instructive to review the effect of systematic manipulations of each factor on vocal fold oscillation.

Airflow Pressure and the Bernoulli Effect

Vocal folds oscillate by a combination of the redundancy of the vocal fold mucosa and the Bernoulli effect. The Bernoulli effect states that as flow is pushed through a constricted narrowing such as the glottis, the pressure across the narrowed orifice will be inversely proportional to the velocity of the flow. The same volume of air must flow through a constriction; therefore, there is a rapid rise in the velocity profile at the glottis. With this constriction, the pressure and velocity profile of the flow will change. Downstream from the trachea, the flow is under great pressure to move, but the wide tracheal diameter converging into a constricted glottis results in a mean velocity of flow that is low at the trachea and high at the glottis. At the site of constriction of vocal folds, the narrowing in the vocal folds means that the same flow across the trachea must traverse the narrowed vocal folds at a greater velocity so as to maintain the same flow rate. The velocity of airflow across the adducted vocal folds is much greater than in the trachea. Flow, which passes through a constricted orifice at high velocity, will create a negative transmural pressure gradient in proportion to the velocity of flow. Thus, at the narrow vocal folds, there is a negative transmural pressure acting on tissue making up the vocal tract. The effect of this negative transmural pressure relative to high velocity of flow across a constriction is called the Bernoulli effect. The Bernoulli effect will act to counter the positive subglottic driving pressure by a negative pressure. This negative pressure will tend to collapse where the constriction is greatest. By this theory, the faster a flow is directed through a tube with a constriction, the greater the collapsing forces will be exerted at the most constricted part of the tube. With the vocal folds being elastic and deformable, the flow across the vocal folds will collapse the vocal folds by sucking them inward until flow has ceased. When flow has ceased, the velocity of flow across the orifice falls to zero, and the negative transmural pressure falls to zero. The buildup of pressure downstream then becomes the dominant force now acting to open the vocal tract by pushing the passive vocal folds apart (**Fig. 45–3**). The vocal folds are pushed open by subglottic pressure until flow is reestablished across the vocal folds. At this point, the cycle of Bernoulli effect and flow-related collapse is repeated. With the balance of steady subglottic pressure and pliable focal folds, this continuous oscillation may be maintained at a quasiperiodic oscillatory state. This steady oscillation will continue until (1) the cessation of breath or loss in breath support, (2) the opening and abduction of the vocal folds, and (3) the release of vocal fold tension. One can see that according to the aerodynamic myoelastic theory of phonation, the actual determination of the vocal tract configuration and tension is under voluntary control, but the vibration of the vocal folds is passive and is caused by a combination of subglottal air pressure and the Bernoulli effect. We will next consider the factors of tissue mass, stiffness, and elasticity as they affect oscillation of vocal folds.

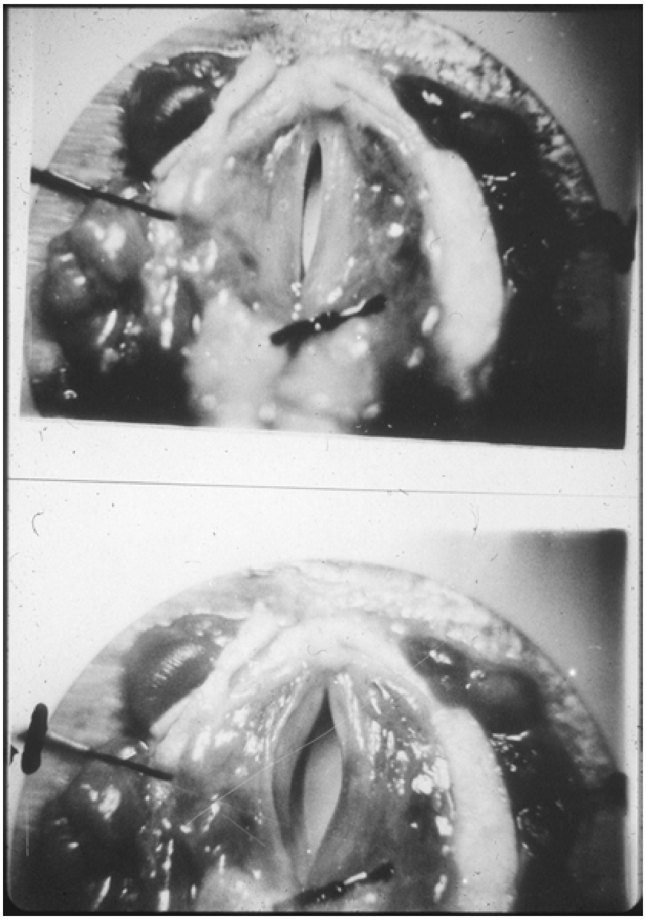


Figure 45–3 *The difference in pulmonary driving pressure on vocal fold vibration is demonstrated on this excised larynx model. The upper photo shows the larynx vibrating at 8 cm of subglottic pressure and ~200 cc/sec. The lower photo shows the same larynx vibrating at 18 cm subglottic pressure and 450 cc/sec flow rate. The sound generated was louder from the lower model. Notice the greater amplitude, wave, and phase difference in the lower photo.*

Stiffness and elasticity are fundamental characteristics of membranous oscillatory bodies made of biological tissues. The vocal folds have a stiffness that tends to resist deformation by pressure. Once deformed, the vocal folds are said to have elastic recoil. This elastic recoil is a restorative force, which will try to restore the deformed vocal fold back to the original configuration. Thus vocal folds are described as having two properties: stiffness, which is the resistance of the tissue to undergo deformation, and tension, which describes the elastic recoil forces that will try to restore the vocal folds once they have been deformed. Because the vocal folds are capable of great change in length, tension, and stiffness, a variety of fundamental frequencies at different loudness levels may be produced with the normal vocal oscillator.

The oscillatory trajectory of vocal folds describes an elliptical pattern. The mass and the length of the vocal folds determine the frequency of vocal fold oscillation. Longer vocal folds will trace out a greater ellipse with greater amplitude of excursion and lower frequency. Thus longer vocal folds under the same tension will describe a larger ellipse, greater displacement, and a lower frequency. Vocal folds, which are of greater mass, will have greater inertia to vibration. Greater force will be necessary to set the folds into oscillation for a heavier mass of vocal folds than a thinner, lighter mass. The greater mass of the vocal folds will have greater inertia and require greater effort to set into oscillation. The more massive vocal folds will also oscillate at lower frequency than vocal folds of lower mass and will require greater driving pressure.

Vocal folds are not uniform in compliance and mass. The midmembranous vocal fold has the greatest compliance of the vocal folds. At the anterior attachment of the vocal ligament and the posterior attachment of the vocal process, the vocal ligament is extremely close to the surface epithelium. The layered structure of the vocal folds loses its separation at the anterior and posterior macular lava. Normally, vocal folds are of equal length, tension, and mass, and paired vocal folds will oscillate in phase and at the same frequency. Thus paired oscillators of the same biological property will oscillate with similar amplitudes and frequencies. If the vocal folds are similar in tissue properties, the paired oscillators will be subject to the same pressures and will have similar equivalent mass, length, tension, compliance, and elasticity. Thus the paired oscillators will act in synchrony and have similar configuration of oscillation, which will in general mirror each other. Any disturbance in tissue rheology, stiffness, and mass will demonstrate an appreciable difference in symmetry, amplitude, frequency and periodicity.

VOCAL FOLD VIBRATION

We have considered how pressure and elastic forces interact to create stable self-sustained vibrations. Critical in the consideration of the oscillation of the vocal folds is the soft, pliable mucosa of the vocal folds. The more pliable and mobile the mucosa, the more the mucosa will move under the influence of the myoelastic aerodynamic effect throughout the glottal vibratory cycle. Great mobility of mucosa, therefore, is one of the critical conditions for ease of voice onset and sustained voice production. Besides great pliability of mucosa, the mucosa must have freedom of movement. Freedom of movement of mucosa must therefore translate into some redundancy of mucosa. At the vocal folds, tissue mobility must take into account the stiffness and redundancy of the mucosa. The redundancy of mucosa and the pliability of vocal fold mucosa will be a critical factor in the consideration of phonomicrosurgery procedures. One final factor in the consideration of ease of vocal fold vibration is the viscosity and tack of the mucosa at the air to mucosal interface. Air passing over wet mucosal surfaces will tend to create a more laminar flow profile than dry air passing over dry mucosa. Dry air passing over irregular dry surfaces will create eddies and vortices of turbulent flow. Turbulent flow profiles across the vocal folds will increase glottal resistance and increase the work of phonation. This is an important though not often studied factor in phonatory physiology. This factor has been termed rheology of the mucous membrane by Isshiki (1989). We can appreciate the importance of moist mucosa in production of sound by the difficulty of speaking when one is nervous and has a dry mouth.

We can appreciate the great variability of normal vocal fold oscillation in high-speed film or stroboscopic video examinations of the vocal folds. Vocal folds are capable of oscillation with and without contact with the contralateral fold. The vocal folds are capable of vibration at great frequencies via changes in thickness and length. The following description details the vocal fold vibratory patterns in three basic considerations: register, loudness, and frequency.

Registers refer to the vibratory patterns of the vocal folds seen during the glottal cycle and are characteristic of the pattern of vocal fold vibration. The three registers are the chest or modal register, the falsetto or loft register, and the vocal fry or pulse register. **Fig. 45–4** depicts the vibratory pattern of the vocal folds for one glottal cycle produced in chest voice. **Fig. 45–5** shows the vibratory pattern of the vocal folds produced in falsetto voice. Vocal fold vibration in the modal or chest register is demonstrated by a glottal cycle pattern with distinct open and closed phases. The vocal folds during the open phase will have an opening phase; each fold will reach maximal amplitude of lateral excursion, followed by a closing phase. The open phase will be followed by a closed phase of vibration. The open and closed portion will be of approximately the same duration. During the closed phase, the vocal folds of the opposite vocal fold are in contact with the contralateral vocal fold. The thickness of vocal fold contact will depend in some measure on the compliance

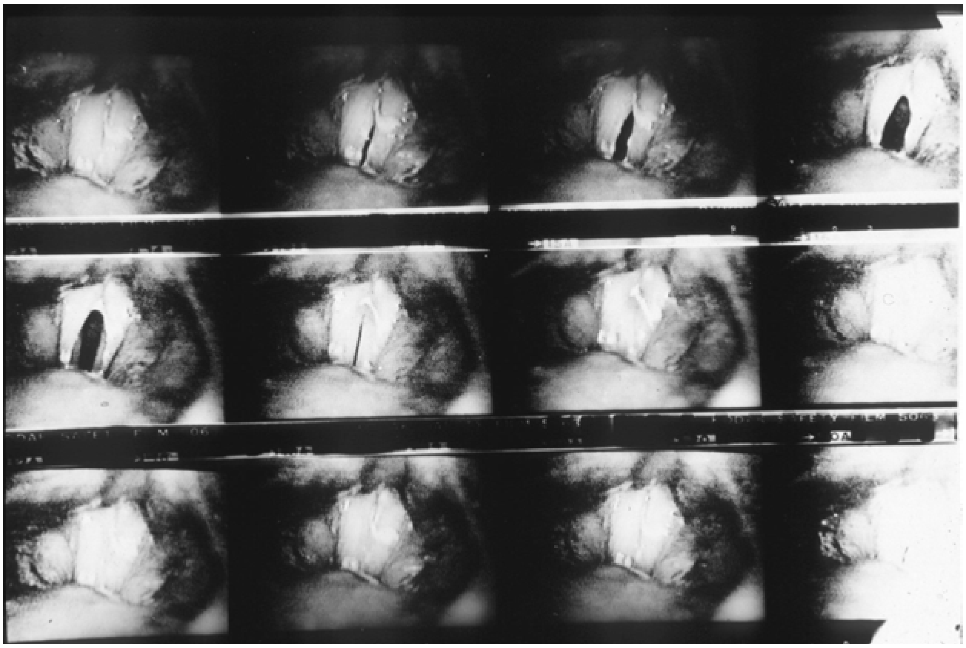


Figure 45–4 *Chest register phonation. The difference in glottal configuration between a falsetto and a modal register is shown in this and the next figure. Notice the well-defined open and closed phases and the amplitude of lateral excursion of each vocal fold in the chest register.*

of the vocal fold tissues brought into contact by the subglottic pressure and the Bernoulli effect. Chest register will be produced by a balance of both thyroarytenoid and cricothyroid muscle activity. The tension on the vocal folds will be low, and the vibration of the vocal folds will show the greatest amplitude compared with the other registers. In general, the greater compliance of the vocal folds seen during chest register will produce a vibratory frequency lower than a noncompliant vocal fold placed under great tension, such as during falsetto (**Fig. 45–5**). The compliant nature of the vocal folds will also be able to shape the acoustic wave and the intensity of the acoustic wave. When the vocal folds oscillate in phase, each vocal fold will participate in the sudden compression and rarefaction of the airstream simultaneously. This simultaneous action will contribute to the complexity of vibratory motion, thereby amplifying the effect of vocal folds and producing a louder, crisper acoustic quality. The acoustic wave signature will have greater signal and

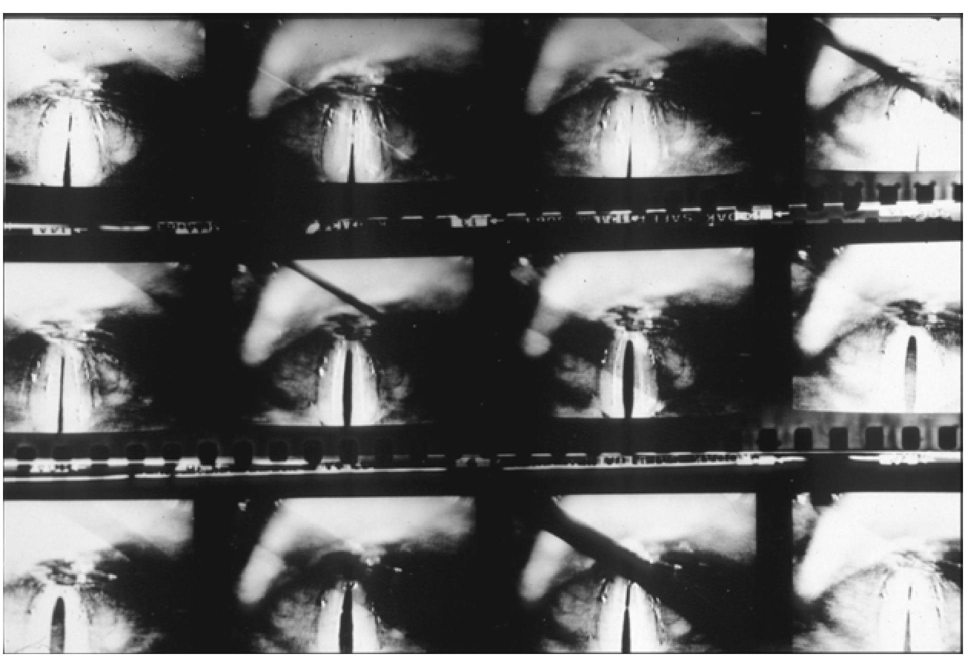


Figure 45–5 *Falsetto register phonation. The vocal fold is tenser, longer, and vibrates without an open and closed phase. The amplitude of vocal fold vibration is small, and the mucosal wave is not visible.*

frequency spectra than if only one vocal fold was able to participate in vibration.

Vibration of the vocal folds has an elliptical trajectory. Its movement may be described using a skipping rope analogy. The glottic opening starts at the lower margin or lip of the vocal folds by deformation of the tissue. This is the initial mucosal upheaval. The mucosal upheaval is propagated cephalad as a disturbance of the mucosal wave moving across the rounded surface of the vocal fold. As the mucosal deformation traverses the vocal fold edge, it will go laterally across the superior surface of the vocal folds. This can be observed readily by stroboscopy. The mucosal upheaval and the resultant mucosal wave are thought to have great significance in determination of the acoustic wave generated at the vocal folds.

The trajectory of vocal fold oscillation is elliptical. This elliptical nature of oscillation is demonstrated by the opening at the lower lip of the vocal folds prior to the upper lip. This is the normal vertical phase difference between the upper and lower lip of the vocal fold. With the upper lip opening at its maximum, the lower lip is already closing. This phase difference between the upper and lower lip is exaggerated when the voice is made louder. The normal vertical phase difference is indicative of pliable mucosa and is important clinically. Patients suffering from dysphonia may have loss in the normal vertical phase difference between the upper and lower lip of the vocal folds. Lesions of the lower lip of the vocal fold may be invisible to the examiner unless the vertical phase lag is exaggerated by having the patient phonate loudly. This maneuver brings the lower lip into view while the upper lip is still opening.

The unique layered structure of the human vocal fold is important to its oscillatory capabilities. Many authors, including Hirano (1981) and Gray, have demonstrated the histologic properties of the vocal fold. The histologic demonstration of the layered structures has shown that the mucosal epithelium rides on a loose superficial and intermediate layer of the lamina propria. The loose lamina propria unique in the human vocal fold is demonstrated by injection in **Fig. 45–6**. This layer is capable of great deformation and recovery associated with mechanical oscillation of the vocal folds. When there is inflammation or scarring, the layered viscoelastic structure is often destroyed or stiffened. The increased stiffness of the vocal fold cover will then be difficult to set into vibration. This will change the vocal oscillation in several ways. Probably the most sensitive to change is the propagation of the mucosal wave. Another feature of loss of the normal vibratory structure is the loss of the normal phase lag between the upper and lower lip of the vocal folds. With progressive stiffness, the vibratory amplitude will be affected and result in a nonvibrating vocal fold edge. The pliable superficial layer of the lamina has been termed the vocal fold cover. The intermediate and deep layer of the lamina propria have more collagen. It serves as the anchor for insertion of the vocalis muscle. The main mass of the vocal fold is the body, which has the thyroarytenoid muscle is its main component. The vocalis muscle, by insertion on the vocal ligament, is able to adjust mass and tension of the vocal fold. The body does not have the fluidlike viscoelastic properties of the vocal fold cover and therefore exhibits less ability to deform and oscillate. The vibration of the vocal folds occurs mainly on the vocal fold cover and rides on the body of the vocal fold. This has been called the body cover theory of vocal fold vibration.

Figure 45–6 *Human larynx vocal fold injected with ink into the lamina propria shows limited uptake in the membranous folds.*

ACOUSTIC CHARACTER OF THE GLOTTAL SOUND

The oscillations produced by the vocal folds have a complex acoustic waveform. The complex tone may be analyzed into frequency components of complex frequency spectra by using Fourier's analysis. Pure tones have spectral energy in only one frequency and can be described by a single sinusoidal sine wave. Complex pure tones are composed of relative contributions from the fundamental frequency of vibration and its harmonics (**Fig. 45–7**). When the vocal folds generate a sound that is no longer periodic or is contaminated by other extraneous sound sources, these sounds are termed noises. Noises are different from

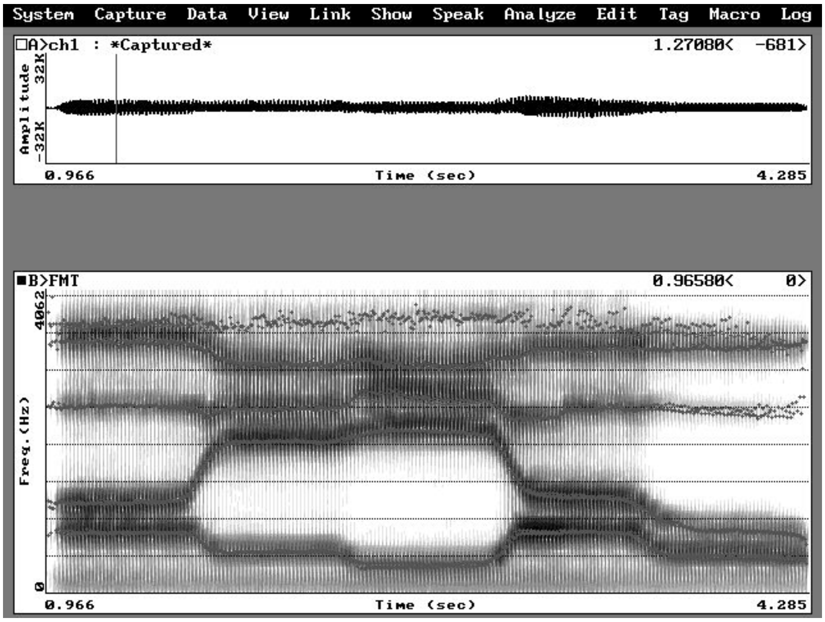


Figure 45–7 *Normal voice signal (top) and spectrogram (bottom) of the vowels a, e, I, o, and u. The spectrogram shows very distinct regions of spectral energy or bands for each vowel. Each band represents an area of intensified energy.*

pure glottal tones by their lack of periodic vibration. Vibration will be lacking in the frequency banding noticeable on analysis of a complex tone. The wide band noise may be located in different frequencies but is characterized by a lack of frequency specification (**Fig. 45–8**). Noises, which are generated in the vocal tract by random airflow, are termed turbulence sounds. When turbulent sound sources combine with the pure tone sounds generated by the vocal folds, there is a mixture of complex pure tones and the wide band, nonharmonic sounds. This results in a "smearing" of the distinct bands of the acoustic spectrum. Acoustically, this results in perceptual qualities of a breathy (**Fig. 45–9**) or rough quality.

RESONANCE AND ARTICULATION

The vocal tract amplifies the glottal sound source. The resonating vocal tract may be conveniently divided into various chambers. These include the supraglottic larynx, hypopharynx, oral pharynx, and oral cavity. The resonance characteristics of the vocal tract may help to amplify certain fundamental frequencies and dampen others. The vocal tract is a variable resonator.

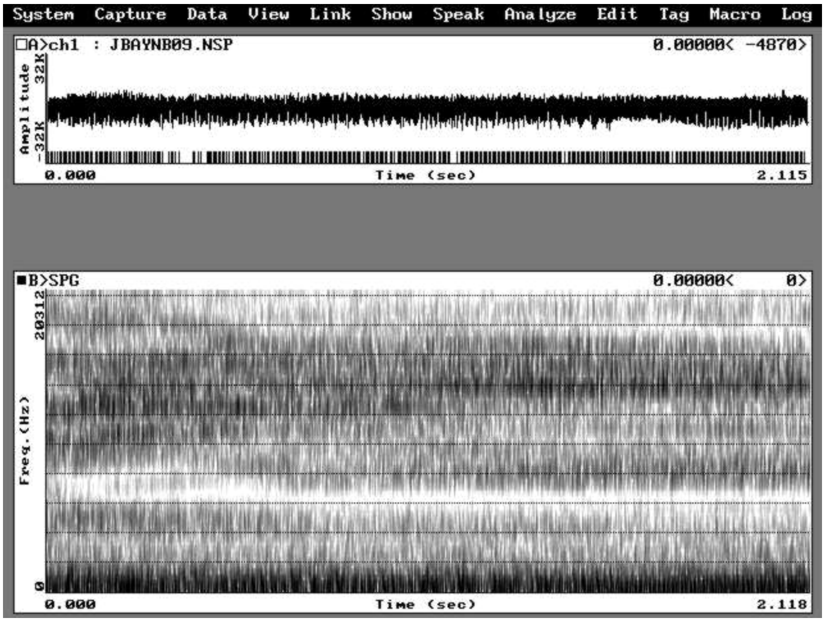


Figure 45–8 *Spectrogram of a hoarse patient, with poor spectral band marking and indistinct spectral bands due to a mass on the vocal folds (papilloma).*

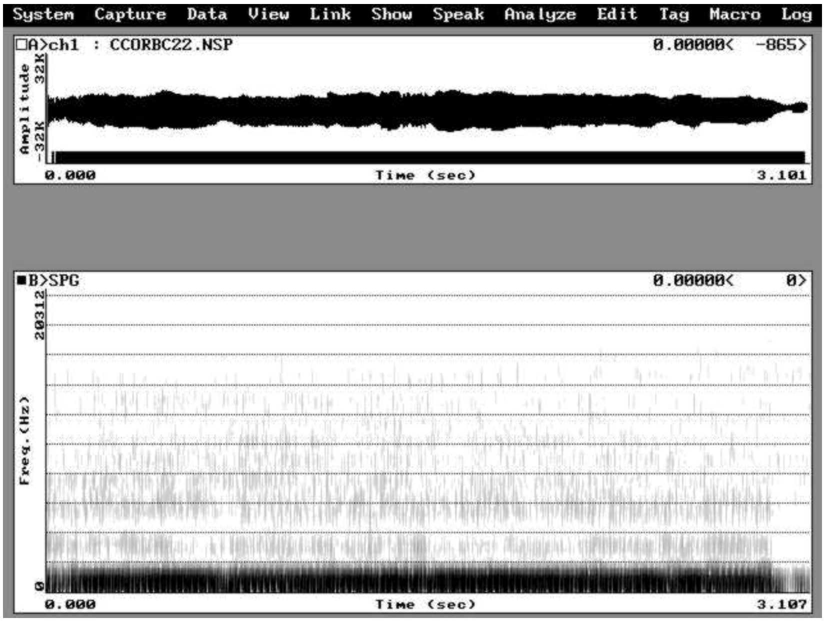


Figure 45–9 *Spectrogram of breathy phonation. There is loss of high-frequency energy. This voice was characterized as breathy and asthenic.*

By its great variation in size, tissue type, and compliance, shifts in resonating characteristics can be used to adjust rapidly to the changes in the fundamental frequency to amplify the sound. Because the resonator is soft and has great configuration changes, there are multiple possible frequency amplifications. This makes the resonator a rather flexible resonator but not a sharp amplifier. As a general rule, the smaller resonators will give higher frequencies better amplification. The vocal tract has many resonant frequencies depending on the laryngeal height and the tongue, lip, and palatal shape. One example of how resonators change the glottal sound source is by the production of different vowels. The placement of the tongue, lips, pharynx, and palate will shaped the pharyngeal and oral cavity to amplify selectively the fundamental frequencies to the specific resonance curve for that vowel or sound quality. These resonant frequencies may be studied on a spectrogram. The spectrogram displays the acoustic energy by distinct bandwidth according to the frequencies. The bandwidths that are selectively amplified are displayed in several distinct regions. These distinct bandwidths are darker and indicate more energy at these frequencies. These distinct resonant frequency peaks are called formant peaks (**Fig. 45–7**). Singers must spend many years perfecting the complex dynamic shaping of their vocal tract to selectively amplify the many fundamental frequencies to carry sound over the sound produced by an orchestra. The resonance characteristic for this ability in singers to project sound above the orchestra has been termed the "singer's formant." This is a selective amplification of the F3 (the third formant) and is centered at ~3000 Hz.

VOICE QUALITY

Perceptual rating of the voice is possible with some useful and reasonably agreed on terminology to describe an abnormal voice. A perceptual rating scale, which is often used, is the GRABS scale, which rates the voice by an overall severity *g*rade, *r*oughness, *a*sthenia, *b*reathiness, and *s*train. The ratings made by trained listeners are 0 = normal, 1 = mild, 2 = moderate, and 3 = extreme. These characteristics of the voice depend on the shape of the glottal sound, the frequency of the glottal sound, the periodicity of the glottal sound, and the amount of aperiodic noise generated by the vocal tract during the production of the voice.

One important aspect of the generation of the glottal sound is the manner and velocity of the glottal opening and closing. For example, a voice sung in chest register will have glottal opening and closing pattern that exhibit distinct open and closed phases. A voice produced in falsetto or head register will exhibit a sinusoidal oscillation of the vocal folds, which does not exhibit distinct open and closed patterns. Rather, the vocal folds will be either in the opening or the closing phase. The patterns of opening and closing of the vocal folds at different registers are shown in **Figs. 45–4** and **45–5**. Another example of why the opening and closing patterns are important is the role they have in regulating sound intensity. This is dependent on a combination of (1) subglottic pressure and glottal power, (2) glottal efficiency, and (3) transfer function of the vocal tract. The driving pressure of vocal fold vibration is a product of the driving pressure built up in the subglottis. This is by the combination of diaphragmatic and thoracic muscles driving against a closed glottis. The product of subglottic pressure and airflow rate is defined as glottal power. With the same glottal resistance, a greater glottal power used to drive the vocal folds will result in greater intensity of sound. This is usually seen as increased effort of exhalation.

A second mechanism to increase intensity is provided by the glottis. With a greater adduction of the vocal folds, the glottal resistance will increase. This will result in a shorter opening phase of the vocal folds, a sharper glottal pulse. As air is pulsed through the glottis in a shorter opening period, the rate of vocal fold opening and closing will be accelerated. The glottal sound pulse produced will be sharpened. The acoustic spectrum of the glottal sound will be more intense at the fundamental frequency and its harmonics. The opening and closing velocity of the vocal folds has great impact on the intensity of the sound produced. From the preceding discussion, we see that opening and closing rates of the vocal folds are affected by both the driving pressure of the lungs and the efficiency of the glottis. The third mechanism of increasing intensity is by the transfer function of the glottal sound through the resonating vocal tract. The different vocal tract shape has a characteristic transfer function, which amplifies certain frequencies and dampens others. If the fundamental frequency and its harmonics are matched to the acoustic transfer function of the vocal tract, then the voice produced will be amplified.

The fundamental frequency of vibration of the vocal folds is the rate of vibration of the vocal folds, which are vibrating to change the airstream from the lungs into distinct pulses. The fundamental frequency of vocal fold vibration will create a sound at that frequency which is perceived as of a certain pitch. When speaking of pitch, the term *octave* is often used. A note one octave higher refers to the frequency twice that of the previously referred to frequency. Thus, in a person with a fundamental frequency of 100 Hz, one octave higher will be 200 Hz. A person with the lowest fundamental frequency of 200 will have at one octave higher a frequency of 400 Hz. The mean frequency of vocal fold vibration for men is lower than for women. During speaking, male larynges vibrate between 100 and 200 Hz during speech, while female larynges are usually about one octave higher (200–400 Hz). The larynx is capable of much greater ranges of vocal vibration during singing. Frequencies from a low of 60 Hz can be produced by bass singers, and fundamental frequencies greater than 1000 Hz can be produced by soprano singers. Typically, a normal adult will have a 2- to 2½-octave useful range. This can be tested by a phonetogram.

How vocal folds regulate fundamental frequency of vibration is changed by three factors that are closely intertwined: vocal fold length, mass, and tension. Although vocal fold length and mass are determined by gender, age, and natural development, the main adjustment of these factors in daily speech is by the changes brought about by the intrinsic and extrinsic laryngeal muscles of the larynx. Increasing the tension on the vocal folds will allow a shorter, thinner portion of the vocal folds to vibrate. This will increase the frequency. A major increase in vocal fold tension will occur if the vocal folds are stretched. This action is by the cricothyroid muscle. Thus the cricothyroid muscle will elongate the vocal fold, increase the tension of the vocal fold, and increase the stiffness of the vocal fold. Thinning the vocal folds will also change the mass of the vocal fold. The thinner, lesser mass of the vocal folds results in vibration at a higher fundamental frequency of vibration than a larger, massive vocal fold. The mass of the vocal folds is longer and thicker in males and results in a lower fundamental frequency, much as the strings on a viola are different than a violin's. Vocal folds may be changed in mass by the selective tension of the vocalis and the cricothyroid muscle. When the cricothyroid muscle is relaxed and the thyroarytenoid muscle is active, the vocal folds will shorten and become thicker. The thicker vocal folds will have a greater mass and vibrate at a lower fundamental frequency, resulting in frequency shifts within the same person.

When vocal folds are elongated by action of the cricothyoid muscle, there is also an increase in tension and stiffness. This is accompanied by an increase in vocal fold thinning. This increased stiffness and reduced mass more than make up for the increased length of the vocal folds. The net effect of cricothyroid contraction is pitch elevation.

We have examined how the phonatory system is a complex system made of three subsystems. The respiratory driving system provides the power. A phonatory system interacts with the respiratory system to convert airflow from the lungs into a rapidly alternating airflow stream that produces the glottal sound source. The vocal tract amplifies the sound produced at the glottis. A multisystem function produces voice. We will now consider how these systems can go wrong in the pathophysiology of voice production.

PATHOLOGY OF VOICE PRODUCTION

Pathophysiology of Dysphonia

Analysis of the patient with dysphonia starts by general examination of the patient's respiratory, neurological, and articulator systems. If these systems are deemed normal or near normal, then the analysis of the dysphonic patient is then focused on the mechanical phenomenon of phonation production. The critical factors to consider are abnormal subglottic pressure, abnormal glottal configuration, scarring and stiffness of the vocal fold margin, lack of pliability of the mucosa to permit sliding of the vocal cover on the body, and alteration in the mass or tension of the vocal folds. These will be considered in detail.

Abnormalities of the subglottic pressure without intrinsic laryngeal abnormality are commonly attributed to neurological, muscular, or functional disorders. Examples of these include soft voice in patients with Parkinson's disease, myasthenia gravis, and dysphonia of conversion disorder. Respiratory diseases such as chronic obstructive pulmonary disease (COPD) can result in reduced pulmonary capacity and subglottic pressure.

Patients with dysphonia often have incomplete closure of the vocal folds. This can be best detected by stroboscopy or laryngoscopy. Imperfect closure of the vocal folds may occur due to tumor, benign growths, or vocal cord paralysis or vocal atrophy (**Fig. 45–10**). If a large airflow is leaking across the larynx, the sound will be soft and breathy in character. Air leakage across an orifice will also result in turbulent noise generated by nonmodulated flow. Turbulent flow is characterized by high-frequency random energy. This results in acoustic spectrograms that are smeared and lacking in discrete formants. The perceptual result from this is roughness or breathy voice quality. The imperfect approximation of the vocal folds will be lacking in ability to oscillate periodically. Imperfect approximation of the vocal fold will fail to produce a glottal sound source crisp and rich in harmonics. Hoarse voices will be composed of aperiodic, random noise from turbulence of airflow. Imperfect closure of the vocal folds may be secondary or primary in etiology. Examples of primary etiology are vocal cord paralysis with big gap and prebyphonia with bowing. Secondary imperfections of vocal fold closing may be due to mass lesions of the folds that prevent closure of the vocal folds.

During phonation, the glottal tension is regulated so that it will be tight enough for the vocal folds to oscillate and create relatively complete closure. The size of the glottal gap is minimal. The pre-phonatory set of the vocal folds will result in a small gap that is just large enough to be closed by the oscillation of the vocal folds. Too large a gap during prephonatory set will result in breathy phonation, whereas too tight a glottal closure will result in harsh or strained quality of voice. If a high subglottic pressure is necessary to set the vocal folds into oscillation, the voice will sound strained and pressured.

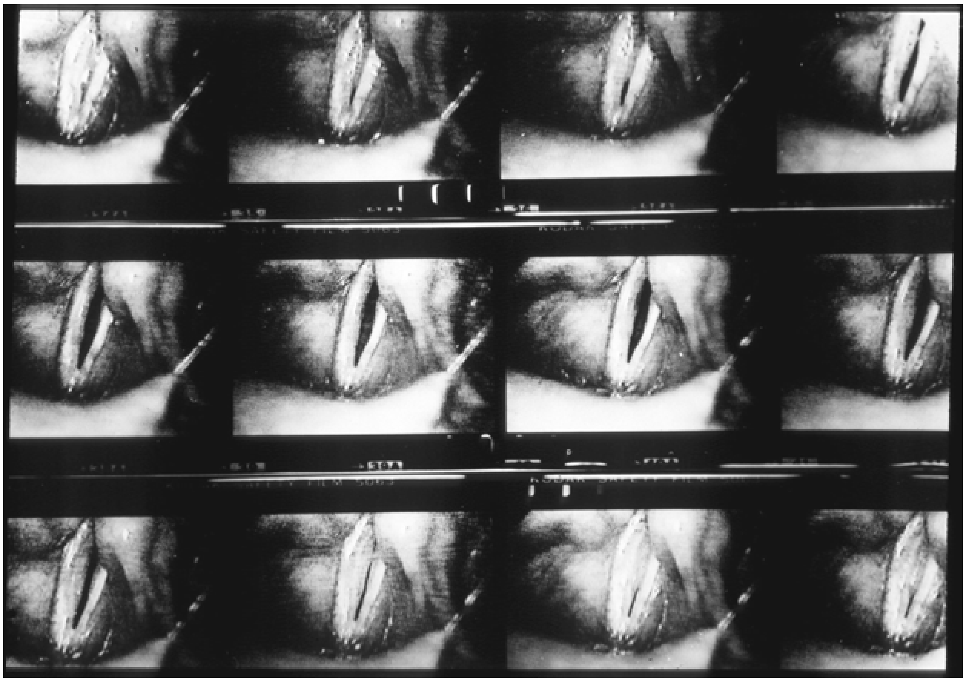


Figure 45–10 *Mass and tension differences can result in dysphonia. The left vocal fold (right) is paretic and has less tension. The amplitude of vibration is less, resulting in a glottal gap.*

Stiffness and mass abnormalities of the vocal fold mucosa are common causes of dysphonia. When masses of the vocal folds are too great, the masses may be unable to oscillate without excessive subglottic pressure. Too little mass of the vocal fold seen in patients after vocal fold stripping often results in vocal folds that are too stiff or have too little compliance to vibrate. Scarring of the vocal folds after surgery is a common cause of voice difficulty. The scar is stiff and lacks the soft superficial layer of the lamina propria. Stiff vocal folds take excessive subglottic pressure to set into vibration. Lack of compliance of the vocal folds results in small vibratory amplitude of the mucosa. There is often a lack of mucosal wave. The loss of amplitude and mucosal wave in a scarred or stiff vocal fold can be seen as an asymmetric vibration compared with the normal vocal mucosa. This can be detected with stroboscopy.

Imbalance of the masses or tension of the vocal folds is another source of abnormal voice. An imbalance of vocal fold tension may come about in patients with vocal fold paresis and in patients with unilateral edema or stiffness of the vocal folds. When this occurs, vibration of the vocal folds will have a vocal fold that leads the other fold during vocal fold opening. This results in a glottal vibratory cycle that has a phase shift between the folds. The net result of a phase lag between the vocal folds is a loss in efficiency in the glottal flow converter. Such a loss in flow conversion and efficiency may result in subtle yet significant dysphonia.

ANATOMY OF THE VOCAL FOLDS AND SKELETON RELEVANT TO PHONOSURGERY

Isshiki (1989) initially designated four different types of laryngeal framework surgery that were possible for different clinical conditions. Type I thyroplasty is a form of medialization laryngoplasty where an implant is placed between the thyroid cartilage and the vocalis muscle to medialize the membranous vocal fold. This is the most popular type of laryngeal framework surgery because it addresses the most common clinical disorder of glottic incompetence due to vocal fold paralysis. Arytenoid rotation technique or the arytenoid adduction procedure is a supplemental procedure within medialization laryngoplasty designed to correct the posterior chink and the arytenoid malrotation that are often present in patients with vocal fold paralysis. Type II thyroplasty is lateralization laryngoplasty. This involves the lateralization of the thyroid cartilage. This creates a greater glottic lumen. The resultant voice would be soft and breathy. This procedure was proposed as a treatment alternative for spasmodic dysphonia. Type III thyroplasty is anteroposterior thyroid cartilage shortening. This would result in a lax vocal fold and pitch lowering. This was proposed as the treatment alternative for puberphonia in postpubescent males with inappropriate high pitch. Type IV is anteroposterior elongation by cricothyroid approximation. The cricothyroid approximation would result in a longer, thinner vocal fold and a higher pitch. This would be appropriate for patients undergoing gender reassignment and those with cricothyroid muscle paralysis. The type I thyroplasty is also termed medialization laryngoplasty. Because it is performed most frequently for medialization of the vocal folds in patients with unilateral vocal fold paralysis, medialization laryngoplasty is the procedure performed most often alone or in conjunction with arytenoid adduction.

Precise knowledge of the level of the vocal fold as projected on the external thyroid cartilage is of critical importance for the performance of thyroplasty type I and all types of laryngoplasty procedures. The key landmarks are projections of the internal laryngeal landmarks on the thyroid cartilage framework. These landmarks are guides for surgery of the laryngeal skeleton. Projection of landmarks on the laryngeal skeleton is used to find key structures in the larynx. These are the mark of anterior commissure, the top of the vocal fold, the muscular process of the arytenoid cartilage, and the posterior cricoid plate.

The distance of the anterior commissure tendon on the thyroid cartilage may be made relative to the most prominent landmark of the thyroid skeleton, the thyroid notch. Development of the larynx in males and females occurs largely on the superior portion of the thyroid cartilage. This makes projection of the anterior commissure on the thyroid cartilage between the lower border of the thyroid notch and the lower border of the thyroid cartilage reasonably constant. Half the distance between the lower border of the thyroid cartilage and the lower border of the thyroid cartilage will point to insertion of the anterior commissure tendon. For males, this will be between 10 and 12 mm from the lower border of the thyroid cartilage, while females will have a point from 8 to 10 mm from the lower border to the thyroid cartilage. This will give an accurate placement of the anterior commissure in the majority of cases.

Another important line to identify is the projection of the upper surface of the vocal cord. Medialization of the true vocal fold must take place without excessive medialization of the false vocal fold. Placement of the implant too superiorly will risk the possibility of implant extrusion through the ventricle. The line of the vocal fold is found to be parallel to the lower border of thyroid cartilage and arises from the insertion of the anterior commissure to intersect the oblique line on a line parallel to the inferior

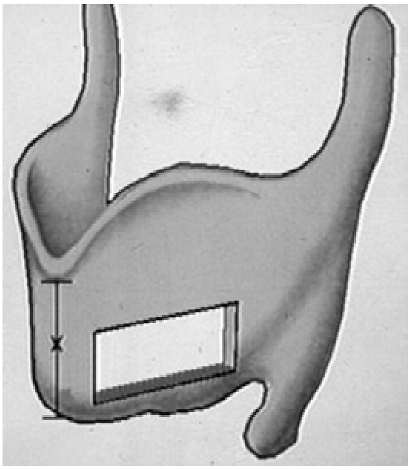


Figure 45–11 *Placement of the thyroplasty window should be parallel to the lower border of the thyroid cartilage, 5 mm posterior to the midline, and on the lower half of the thyroid lamina as measured from the lower border of the thyroid cartilage to the lower border of the thyroid notch.*

border of the thyroid cartilage. Some authors also refer to the line as arising from the anterior commissure to the intersection of the oblique line two-fifths the distance from the bottom of the thyroid cartilage height. From histology of whole sections of the larynx, it is clear that the height of the vocal fold may migrate lower as the vocal fold becomes denervated. The thyroplasty window therefore should always be situated on the lower half of the thyroid cartilage below the halfway mark between the thyroid notch and the lower border of the thyroid cartilage (**Fig. 45–11**). This line is quite different and higher than the line used to mark supraglottic laryngectomy. The purpose of the upper line for the vocal folds is to identify the level above which medialization implants should not enter. In this way, the pitfalls of an implant malposition and extrusion can be avoided.

Identification of the muscular process of the arytenoid cartilage is necessary during the arytenoid adduction procedure (**Fig. 45–12**). The surface projection of the muscular process on the thyroid cartilage is almost an extension of the projection of the vocal cord. The marking on the thyroid cartilage is at the juncture of the lower two-third and upper three-third point of the thyroid cartilage, with the oblique line as the crossing point. Thus, during arytenoid adduction surgery, some authors uses a high-speed drill to remove the posterior 2 × 1 cm of thyroid cartilage to gain rapid access to the muscular process without the need for disarticulation of the cricothyroid joint.

An alternative method of identification of the muscular process and the slightly inferior cricoarytenoid joint is by the use of landmarks on the cricoid cartilage. This uses the relationship of the cricothyroid joint to the

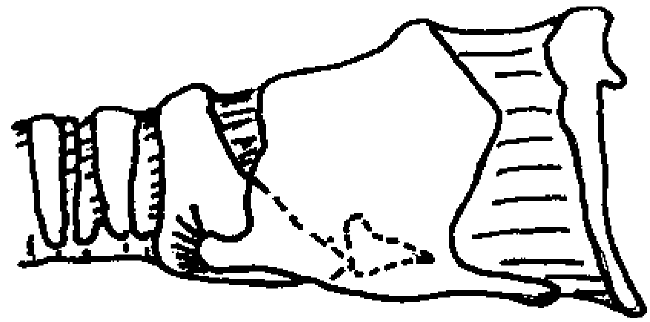


Figure 45–12 *Projection of the muscular process on the thyroid cartilage is represented by the oblique line vertically and the thyroplasty window.*

cricoarytenoid joint. The cricoid cartilage is a signet-shaped ring with a large posterior lamina and a short anterior lamina. On the lateral aspects of the cricoid lamina is the concave, cylindrical cricoarytenoid joint. This joint may have several degrees of freedom. It can roll anteriorly medially and tilt the arytenoid cartilage forward. Or it may rock backward and laterally, producing a rolling of the arytenoid away and tilting the apex of the arytenoid posteriorly relative to the cricoid cartilage. Finding the cricoarytenoid joint is occasionally necessary in exploration of patients with ankylosis of this joint. This joint space is reliably found by disarticulating the cricothyroid joint. From this point the cricoarytenoid joint is less than 1 cm superior and posterior along the cricoid cartilage (0.6–1.0 cm).

The phonosurgeon performing laryngeal framework surgery should have a clear three-dimensional concept of the cartilaginous and soft tissue structures of the larynx. The position of the vocal fold, the cricoarytenoid joint, the arytenoid cartilage, and the intrinsic muscles' origins and insertions are all anatomical details that one must master to become a proficient phonosurgeon.

PHONOMICROSURGERY

The understanding of the uniquely layered structure of the vocal fold was studied extensively by Hirano (1981) and more recently by Gray and coauthors. The basis of phonomicrosurgery of the larynx is a detailed understanding of the anatomy of this structure and its effect on vibratory capability. Understanding the histologic structure of the normal vocal fold is the basis for phonomicrosurgery.

At the glottis opening, the membranous and cartilaginous portions of the glottis can be clearly separated. Only the membranous portion of the vocal fold can vibrate. This soft tissue component consists of the anterior two thirds of the glottis in the adult, and the

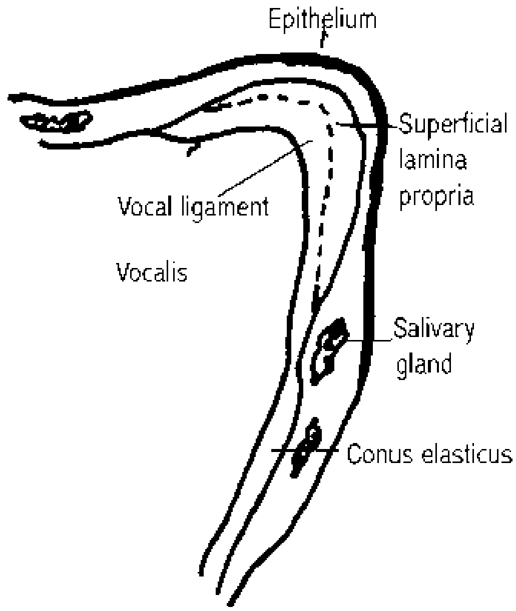


Figure 45–13 *Schematic diagram of the different layers of the vocal fold in coronal section.*

posterior one-third cartilaginous portion is made up largely of arytenoid and cricoid cartilage. The vocal process to the posteriormost opening of the glottis is the cartilaginous portion of the glottis. This space must be closed during phonation while the membranous portion is approximated just enough to permit vocal fold oscillation.

The vocal fold is made up of the vocal fold epithelium, the lamina propria, and the muscular portion of the vocal fold (**Fig. 45–13**). The vocal fold edge is the most important because it relates to vibrator capability. In the human vocal fold, the layered structure of the vocal fold is unique. The mucosa of the vocal fold is stratified squamous epithelium at the vocal fold edge and respiratory epithelium at the superior and inferior surfaces of the vocal fold. The lamina propria is a continuous layer of varying densities. The lamina propria may be separated into layers according to the ratio of collagen and elastin. The superficial layer of the lamina propria is very loose, with different degrees of collagen fibers. The collagen fibers anchor the mucosa to the deeper layers of the lamina propria. The lamina propria layer is also referred to as the Reinke's space. The superficial layer of the lamina propria is very pliable. Its main constituents are water and glycoprotein.

The intermediate layer has elastic fibers that blend into the deep layer. The deep layer is much thicker and made of collagen fibers arranged in parallel. This has also been called the vocal ligament. Into the vocal ligament on the deep surface of the lamina propria are the muscle fibers of the vocalis muscle. The vocal fold, by its gradual transition in density and structure, has great pliability on the edge while anchoring this liable fold to the muscle and ligaments of the larynx. This layered structure is not duplicated beyond the vocal fold edge. Both superior and inferior to the vocal fold edge, there are minor salivary glands that serve to secrete saliva to assist in vocal fold lubrication. The superior surface of the fold consists of respiratory epithelium, minor salivary glands, and the bulk of the thyroarytenoid muscle. On the inferior surface of the vocal fold, the vocal ligament blends into the conus elasticus. The fibers of the conus elasticus insert directly on the mucosa of the subglottic larynx.

Anchoring the vocal folds into place is the anterior commissure tendon and the posterior macula flava, with its insertion on the vocal process of the arytenoid cartilage. These structures are massive collagen and elastic fibers that are a continuation of the vocal ligament. These insert onto the anterior commissure tendon; the vocal process makes this area much less rich in the superficial layer of the lamina propria. Compared with the midmembranous vocal fold, the anterior and posterior portions of the vocal fold will have a thinner lamina propria; it will have greater stiffness and will vibrate less. The layered structure along the length of the vocal fold allows the vocal fold to oscillate like a skipping rope, anchored anteriorly and posteriorly, but free to vibrate in a superior to inferior and a lateral to medial direction.

SUMMARY

The vocal tract is a highly complex structure in which airflow is transformed from the lungs into discrete puffs of air by the larynx. By aerodynamic and aeroacoustic interaction, the glottal sound is produced. The vocal tract to produce tone then amplifies this glottal sound. In a highly intricate way, the human larynx has evolved in unique anatomical and physiological ways that allow it capabilities not found in other animal models. Communication by voice today may be changed by laryngeal framework surgery and phonomicrosurgery. The surgeon in the planning and execution of phonosurgery must understand the unique physiological and anatomical bases of human phonation.

Suggested Readings

Gray SD, Pignatari SS, Harding P. Morphologic ultrastructure of anchoring fibers in normal vocal fold basement zone. Journal of Voice 1994; 8, 48–52

Hirano M. Clinical examination of voice. In: Arnold GE, Winkel F, Wyke BD, eds. Disorder of Human Communication, vol. 5. Wien and New York: Springer-Verlag; 1981

Isshiki N. Phonosurgery: Theory and Practice. Tokyo: Springer-Verlag; 1989
Tucker H. Surgery for Phonatory Disorders. Edinburgh, London, and New York: Churchill Livingstone
Van der Berg JW. Myoelastic–aerodynamic theory of voice production. Journal of Speech and Hearing Research. 1958; 1: 227–244.
Von Leden H. Fono-cirugia. Acta ORL Iber-Americ. 1971; 22: 291

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. Perceptual evaluation of the hoarse voice by the health professional is based on
   A. Fatigue, phonation time, airflow, loudness
   B. Loudness, stridency, tension, softness
   C. Grade, roughness, asthenia, breathy, strain
   D. Asthenia, hoarseness, volume, pitch
   E. Quality of life survey of the patient
2. The landmarks that best describe the projection of the length of the vocal fold on the thyroid cartilage are
   A. Half the distance from the top to the bottom of the thyroid cartilage and along the oblique line
   B. Half the distance from the lower part of the thyroid notch and the bottom of the thyroid cartilage parallel the lower thyroid margin
   C. Half the distance from the top of the thyroid notch to the bottom of the thyroid cartilage and parallel to the top of the thyroid cartilage
   D. Two-thirds the distance from the bottom of the thyroid cartilage to the top of the thyroid cartilage, bisecting the posterior thyroid lamina
   E. None of the above
3. The acoustic description that best distinguishes the different vowels produced during sustained phonation is
   A. Signal-to-noise ratio
   B. Fundamental frequency
   C. Register
   D. Turbulence energy
   E. Formants

Chapter 46

SURGICAL ANATOMY OF THE PHARYNX AND ESOPHAGUS

DOROTHY FRENZ AND RICHARD V. SMITH

This chapter consists of two parts: the first part provides a comprehensive review of the surgical anatomy of the pharynx; the second part reviews the surgical anatomy of the esophagus. Each section describes the clinical boundaries of each structure (i.e., pharynx and esophagus) and its respective nerve supplies, blood supplies, and lymphatic drainage. The section on the pharynx, describes the physiology of this structure and its fascial planes and spaces. Clinical implications of pharyngeal and esophageal anatomy are discussed.

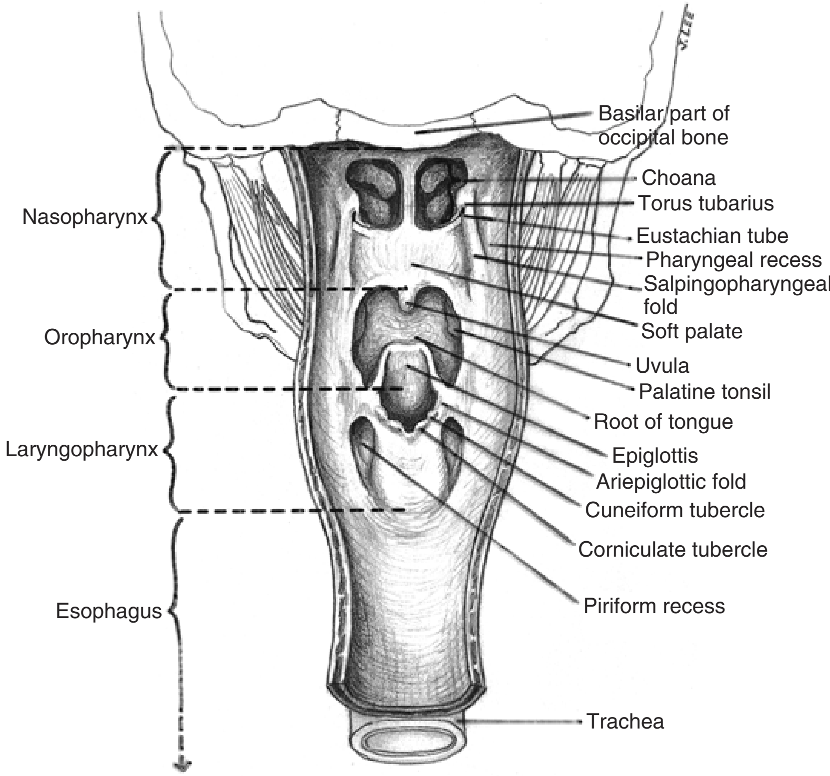


Figure 46–1 *Opened posterior view of the pharynx, demonstrating the boundaries of the nasopharynx, oropharynx, and laryngopharynx.*

STRUCTURAL OVERVIEW OF THE PHARYNX

The pharynx is a fibromuscular, funnel-shaped tube, ~15 cm long, that serves as a common passageway for air and food. It extends from the base of the skull and the styloid process to the inferior border of the cricoid bone. For descriptive purposes, the pharynx is divided into three parts (**Figs. 46–1** and **46–2**): (1) the nasopharynx, posterior to the nasal cavity and superior to the soft palate; (2) the oropharynx, posterior to the oral cavity, between the palate and the hyoid bone; and (3) the laryngopharynx, posterior to the larynx, from the hyoid bone to the inferior border of the cricoid cartilage. The pharynx is widest (~5 cm) opposite the hyoid bone and narrowest (~1.5 cm) at its inferior end, where it is continuous with the esophagus posteriorly. The nasopharynx receives the choanae (i.e., the opening from the nose) and the oropharynx. The posterior wall of the pharynx lies against the prevertebral fascia, with the retropharyngeal space between them (**Fig. 46–3**).

CLINICAL BOUNDARIES OF THE PHARYNX

Nasopharynx

The nasopharynx extends from the base of the skull (basilar part of the occipital bone) to the soft palate (**Fig. 46–1**). Except inferiorly, where it is bounded by the soft palate, the nasopharynx has rigid walls, and hence is continually patent under normal conditions. Superiorly, it is formed by the body of the sphenoid bone and the basilar part of the occipital bone (clivus). The bony wall extends as far as the pharyngeal tubercle (**Figs. 46–2** and **46–3**). Below this, the wall is formed by the pharyngobasilar fascia that overlies the anterior arch of the atlas (**Figs. 46–2** and **46–3**). The cervical vertebrae form the posterior boundary of the nasopharynx (**Fig. 46–3**). Inferiorly, the nasopharynx is continuous with the oropharynx. The pharyngeal openings of the eustachian tube are situated on the lateral walls of the nasopharynx, just posterior to the inferior turbinate (**Figs. 46–2** and **46–3**). The eustachian tubes pass laterally through the sinus of Morgagni, a defect located just superior to the upper edge of the superior constrictor muscle. Posterosuperior to the opening of the eustachian tube is the torus tubarius (a cartilaginous lip) and a collection of lymphoid tissue (i.e., tubal tonsil) (**Fig. 46–1**). Extending downward from the torus is the salpingopharyngeal membrane. The salpingopalatine membrane passes downward anterior to the eustachian tube.

A deep pharyngeal recess, known as the fossa of Rosenmüller, is formed at the angle of the nasopharynx between the posterior ridge of the eustachian cartilage and the posterior wall. The foramen lacerum is situated

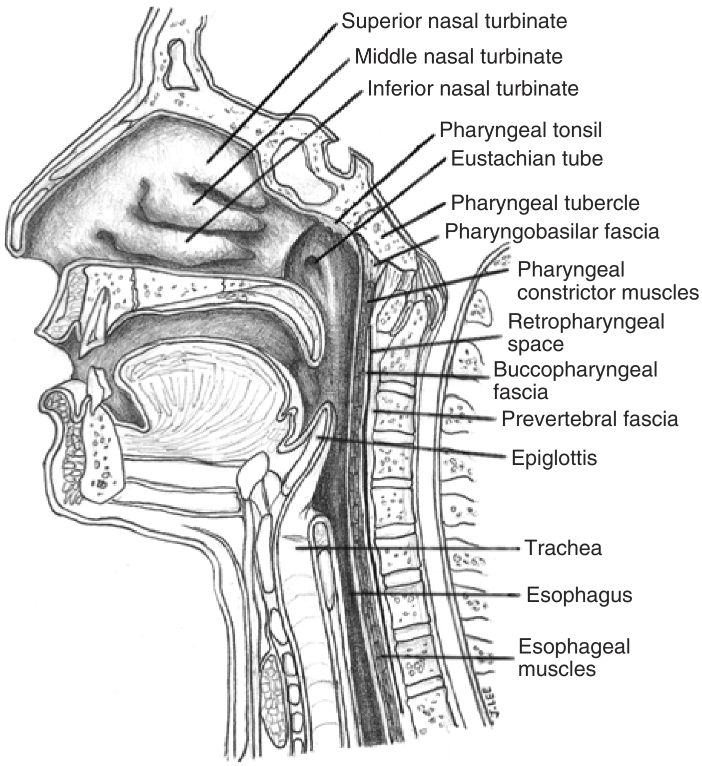


Figure 46–2 *Posterior view of the pharynx: muscles of the pharynx and a partially opened view.*

above this. Because the foramen ovale, foramen spinosum, carotid canal, and jugular canal also lead from the nasopharynx, disease may spread from the nasopharynx to the interior of the skull. Lymphoid tissue (Luschka's tonsil) is found just behind the pharyngeal recess, where the roof of the nasopharynx joins the upper posterior wall. The enlargement of this lymphoid tissue is commonly referred to as the adenoids. Although prominent in children, the adenoids atrophy at puberty and disappear in the aged.

Oropharynx

The oropharynx extends from the plane of the hard palate superiorly to the level of the valleculae (also the level of the hyoid bone). The oropharynx communicates with the nasopharynx above and the laryngopharynx below (**Fig. 46–1**). It is continuous with the oral cavity through the oropharyngeal isthmus; that is, the boundary between the oral cavity and the oropharynx.

In addition to the oropharynx proper, the oropharynx is often regarded for clinical purposes as consisting of the palatine arch (**Fig. 46–4**). The palatine arch consists of the soft palate, anterior tonsillar pillar, and the retromolar trigone; that is, the triangular area over the anterior portion of the ascending ramus of the mandible posterior to the third molar. Although anatomically, this area is situated within the oral cavity, it merges with the anterior tonsillar pillar. Therefore, cancer in this area behaves in a manner resembling that of the oropharynx.

The anterior wall of the oropharynx communicates with the buccal cavity. Below this, the glossoepiglottic area is formed by the base of the tongue. The valleculae, the area bounded by the epiglottis, base of the tongue, and pharyngoepiglottic folds, are situated at the lower part of this anterior wall (**Fig. 46–5**). The median glossoepiglottic fold connects the base of the tongue to the lingual surface of the epiglottis and separates the valleculae into right and left. Each vallecula is bounded laterally by the lateral pharyngoepiglottic fold.

The anterior boundary of the lateral wall of the oropharynx is formed by the palatoglossal fold and underlying palatoglossus muscle (**Fig. 46–4**). The palatopharyngeal fold passes downward from the lower edge of the soft palate and backward to the side wall of the pharynx. The palatine (faucial) tonsil lies in the space between these two folds (tonsillar fossa). The tonsil is oval in shape and demonstrates several crypts on its pharyngeal surface. The intratonsillar cleft, which is situated toward the upper pole of the tonsil, is much deeper than the other crypts and may extend to the deep surface of the tonsil, which is covered by a fibrous capsule containing attached fibers of the palatoglossus

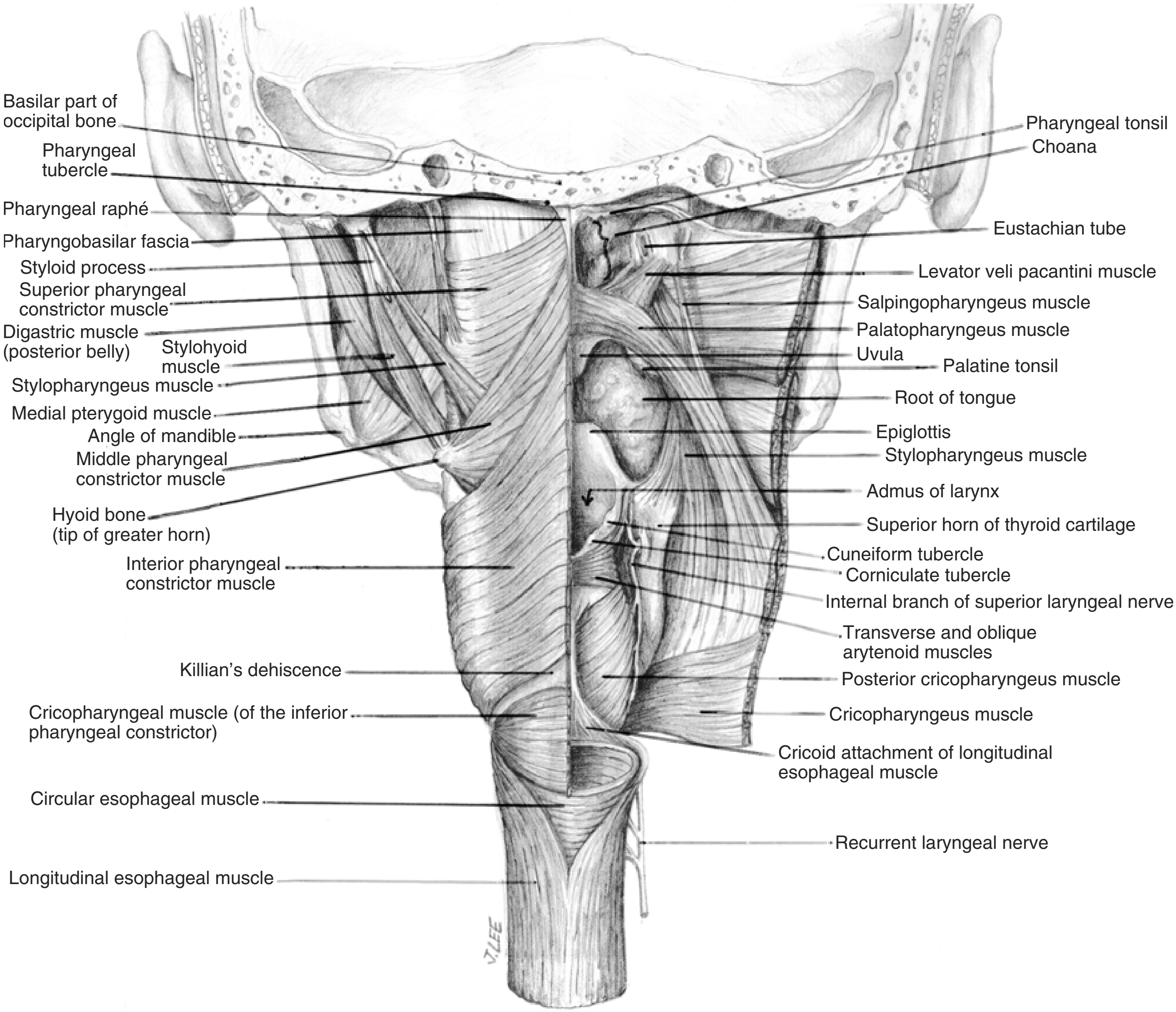


Figure 46–3 *Sagittal section through the pharynx, showing the relationship to the pharyngeal constrictor muscles, buccopharyngeal fascia, retropharyngeal space, and vertebral bodies. A view of the fauces and the pharyngeal opening of the auditory tube is shown.*

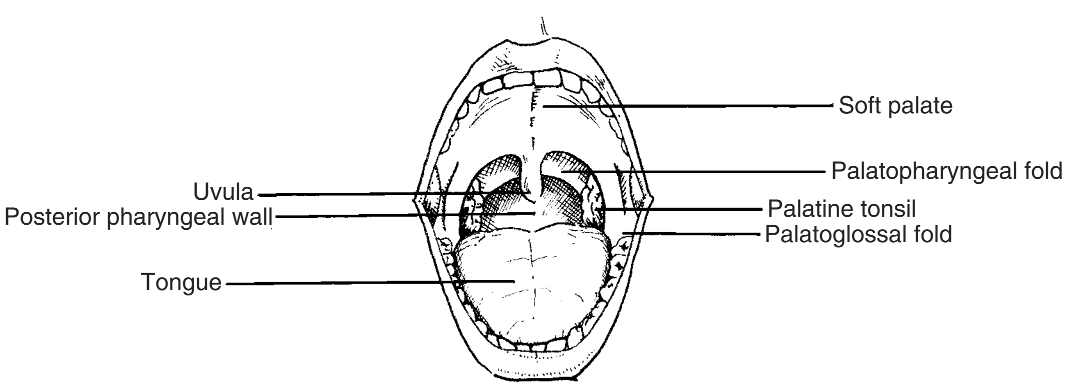


Figure 46–4 *Front view of the mouth and oropharynx, demonstrating the palatopharyngeal and palatoglossal folds. (Modified from Beasley P. Anatomy of the pharynx and esophagus. In: Gleeson M, ed. Scott-Brown's Otolaryngology. Bath, England: University of Bath Press; 1997: Fig. 10.12, courtesy of the publishers.)*

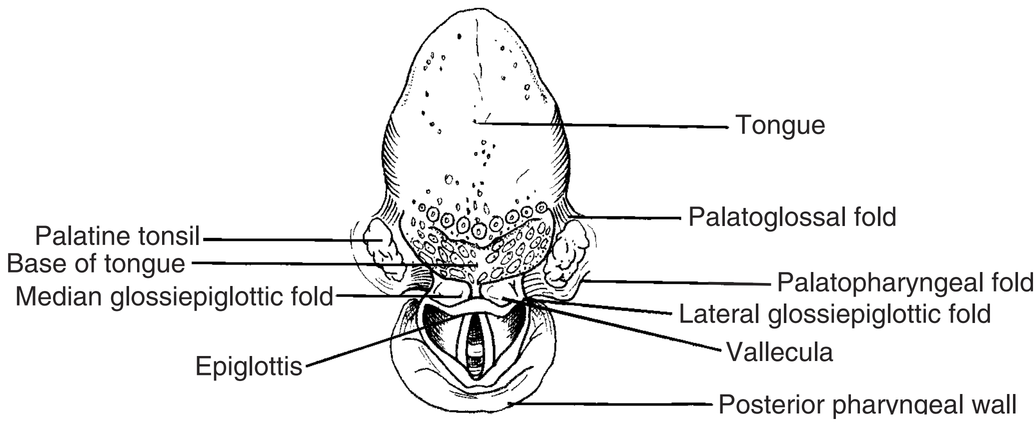


Figure 46–5 *The base of the tongue and valleculae, as viewed from above. (Modified from Beasley P. Anatomy of the pharynx and esophagus. In: Gleeson M, ed. Scott-Brown's Otolaryngology. Bath, England: University of Bath Press; 1997: Fig. 10.13, courtesy of the publishers.)*

and palatopharyngeus muscles. A fold of mucous membrane passes from the palatoglossal fold to the base of the tongue and covers the lower border of the tonsil. In addition, a semilunar fold of mucous membrane extends from the upper part of the palatopharyngeal fold toward the palatoglossal fold. Deep to the tonsillar fossa, the lateral wall of the oropharynx is composed of the superior constrictor muscle and the upper fibers of the middle constrictor muscle. The lateral wall is reinforced by the palatoglossus, palatopharyngeus, salpingopharyngeus, and stylopharyngeus muscles (**Fig. 46–2**).

The posterior wall of the oropharynx is related to the second and third cervical vertebrae and consists of the constrictor muscles, an overlying mucous membrane, and buccopharyngeal and prevertebral fascia (**Fig. 46–3**). The uvula and inferior surfaces of the soft palate make up the superior wall.

Laryngopharynx

The laryngopharynx (hypopharynx) is the region of the pharynx that lies behind the larynx and partly to each side, where it forms the piriform fossae (**Fig. 46–1**). The laryngopharynx is continuous with the oropharynx above and the esophagus below. The larynx lies in the anterior wall of the laryngopharynx. The aditus of the larynx is bounded anteriorly and superiorly by the upper portion of the epiglottis, laterally by the aryepiglottic folds, and posteriorly by the elevations of the arytenoid cartilages (**Fig. 46–2**). Below the aditus of the larynx, the anterior wall of the laryngopharynx is formed by the posterior surfaces of the arytenoid cartilages and the posterior plate of the cricoid cartilage. The pyriform fossae are situated on each side of the larynx and are bounded medially by the lateral surface of the aryepiglottic fold, the arytenoid cartilage, and the cricoid cartilage. They are bounded laterally by the thyroid cartilage. The pyriform fossae extend from the lateral pharyngoepiglottic fold to the upper portion of the esophagus. The superior laryngeal nerve lies deep to a mucosal fold in the lateral wall of the pyriform fossae (**Fig. 46–2**), where it is accessible for administration of local anesthesia.

The posterior wall of the laryngopharynx, which extends from the level of the floor of the valleculae to the level of the cricoarytenoid joint, is formed by the constrictor muscles and overlying mucous membrane. Below this is the pharyngoesophageal junction, or postcricoid area. The postcricoid area is bounded anteriorly by the posterior plate of the cricoid cartilage. It is encircled by the cricopharyngeus muscle (**Fig. 46–2**).

PHYSIOLOGY OF THE PHARYNX

The following section describes the function of the pharynx in digestion, rumination and vomiting, and coordination of craniocervical posture (i.e., maintenance of the airway). The function of the pharynx in swallowing is described in detail in Chapter 47.

Digestion

During swallowing, the pharyngeal glands of von Ebner release a lipase-containing enzyme into the bolus during its passage into the pharynx. These glands are situated below the circumvallate papillae at the junction of the oral and pharyngeal portions of the tongue. In the adult, the openings of von Ebner's glands are opposite the soft palate. During digestion, von Ebner's glands are a primary source of lipase proximal to the pancreas.

Rumination and Vomiting

The pharynx participates in oral emission from the esophagus and stomach. In patients with gastroesophageal reflux, spastic contractions of the esophagus force the esophageal contents upward against the pharyngoesophageal sphincter segment, and potentially, into the pharynx or larynx. This is one pathophysiological mechanism leading to the formation of a pharyngeal Zenker's diverticulum. In patients with esophageal reflux and bolus displacement against the closed pharyngoesophageal segment, hypertrophy of the fibers of the cricopharyngeal sphincter may eventually develop. Intraluminal pressure may rise in the hypopharynx due to cricopharyngeal dysfunction, thereby resulting in outpouching of mucosa in Killian's dehiscence (**Fig. 46–2**).

Coordination of Craniocervical Posture

The pharynx is involved in airway maintenance to the cervical postural musculature. When the mandible is passively lowered during nasal respiration, the tongue and hyoid bone are held forward. However, if there is impairment of airway maintenance, the tongue and hyoid may move posteriorly into the pharyngeal airway, thereby blocking tidal respiration. The competence of pharyngeal airway maintenance can be evaluated by applying pressure posteriorly with the examiner's finger upon the hyoid bone externally or upon the body of the tongue. Suspension of the hyoid bone or tongue in a normal individual should resist this pressure, often with a palpable forward movement against the finger. However, if the person has a weakness of the tongue and hyoid suspensory muscles, the displacement may produce airway occlusion.

THE PHARYNGEAL WALL

The pharyngeal wall is composed of four layers: (1) a mucosa that is composed of ciliated columnar epithelium in the nasopharynx, and a stratified columnar epithelium in the oropharynx and laryngopharynx; (2) a fibrous layer, which forms the pharyngobasilar fascia and is attached superiorly to the skull; (3) a muscular layer composed of inner longitudinal and outer circular parts; and (4) a loose connective tissue layer that forms the buccopharyngeal fascia. The buccopharyngeal fascia is continuous with the epimysium (deep surface) of the pharyngeal muscles and contains the pharyngeal plexus of nerves and veins.

Mucosa

The epithelial lining of the pharynx differs according to its physiological function. The nasopharynx, a passageway for air, is lined by a pseudostratified columnar ciliated epithelium as far as the level of the soft palate (lower border). However, the oropharynx and laryngopharynx, as part of the digestive tract, serve as a passageway for food and are lined by nonkeratinized stratified squamous epithelium. The area bordering the nasopharynx and oropharynx may be lined by a narrow zone of stratified columnar epithelium. Beneath the epithelial lining, there is a lamina propria rich in elastic tissue. Goblet cells are present in the respiratory epithelium of the nasopharynx, and the mucosa is elsewhere pierced by ducts from serous, mucous, or mixed glands that lie within the submucosal layer.

Subepithelial lymphoid tissue encircles the digestive and respiratory openings of the pharynx, forming Waldeyer's ring. Large aggregations of this lymphoid tissue constitute the palatine tonsils in the oropharynx and the nasopharyngeal tonsil (adenoids) in the nasopharynx. Smaller aggregations constitute the tubal tonsil (Gerlach's), the lingual tonsil, and two lateral bands that run posterior to the palatopharyngeal fold.

Fibrous Layer

An intermediate fibrous layer, which forms the pharyngobasilar fascia, is situated between the mucosa and the muscular layers in place of the submucosa. It is attached to the basilar region of the occipital bone and petrous portion of the temporal bone medial to the carotid canal. The pharyngobasilar fascia extends below to the pharyngotympanic tube and forward to the posterior border of the medial pterygoid plate and pterygomandibular raphe. It bridges the gap between the superior border of the superior constrictor muscle and the base of the skull (**Fig. 46–2**). In this region, the pharyngobasilar fascia forms a single layer with the buccopharyngeal fascia. The fascial layer diminishes in thickness as it descends but becomes strengthened posteriorly by a fibrous band that is attached to the pharyngeal tubercle, then passes downward as the midline pharyngeal raphe. The pharyngeal raphe provides attachment to the constrictor muscles (**Fig. 46–2**).

Muscular Layer

An inner longitudinal layer and outer circular layer of muscles contribute to the pharyngeal wall. Each layer is composed of three paired muscles. The stylopharyngeus, palatopharyngeus, and salpingopharyngeus make up the inner layer; the superior, middle, and inferior constrictor muscles make up the outer layer. The constrictors are fan-shaped muscles that arise from the lateral wall of the pharynx and insert posteriorly into the pharyngeal raphe (**Fig. 46–2**). They overlap each other from below upward, forming the side and posterior walls of the pharynx. Intervals formed by the separations at their attachment sites permit structures to pass from the external surface of the pharynx toward its lumen (**Fig. 46–6**). These structures include the palatine branch of the ascending pharyngeal artery, which passes over the upper edge of the superior constrictor muscle; the glossopharyngeal nerve and stylopharyngeus muscle, which enter the pharynx between the middle and superior constrictors; the internal laryngeal nerve, and superior laryngeal vessels, which pierce the thyrohyoid membrane and lie in the piriform fossae; and the recurrent laryngeal nerve and inferior laryngeal artery, which pass between the cricopharyngeal portion of the inferior constrictor muscle and the esophagus. Coordinated contraction of the constrictor muscles during deglutination propels the bolus through the oropharynx into the esophagus. During this movement, the longitudinal muscles elevate the larynx while shortening the pharynx.

Inner Muscular Layer

Stylopharyngeus Muscle

The stylopharyngeus muscle (**Fig. 46–2**) arises from the medial side of the base of the styloid process, descends inferiorly between the external and internal carotid arteries, and enters the pharyngeal wall between the superior and middle constrictor muscles. This long, slender, conical-shaped muscle inserts into the posterior and superior borders of the thyroid cartilage. It elevates the pharynx (and larynx) and dilates the pharynx through lateral expansion, thereby aiding in deglutition.

Palatopharyngeus Muscle

Along with its overlying mucosa, the palatopharyngeus muscle (**Fig. 46–2**) forms the palatopharyngeal arch. It arises as two bundles within the soft palate: the anterior bundle originates from the posterior border of the hard palate and from the palatine aponeurosis, and the posterior bundle arises in contact with the mucosal covering of the palate. The two bundles unite at the posterolateral border of the palate and are joined by fibers of the salpingopharyngeus muscle. The palatopharyngeus muscle passes behind the tonsil forming the posterior tonsillar pillar and inserts into the posterior border of the thyroid cartilage. It elevates the pharynx (and larynx), shortening the pharynx during deglutition and constricting the palatopharyngeal arch.

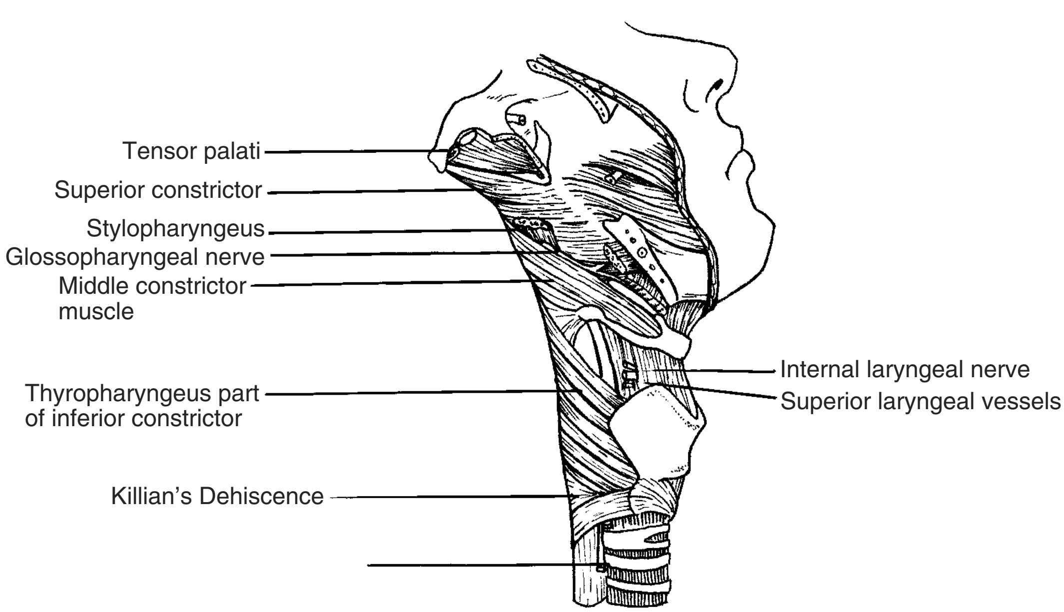


Figure 46–6 *Lateral wall of the pharynx, demonstrating the constrictor muscles and associated structures. (Modified from Beasley P. Anatomy of the pharynx and esophagus. In: Gleeson M, ed. Scott-Brown's Otolaryngology. Bath, England: University of Bath Press; 1997: Fig. 10.12, courtesy of the publishers.)*

Salpingopharyngeus Muscle

The slender salpingopharyngeus muscle (**Fig. 46–2**) arises from the cartilage of the pharyngotympanic tube at its pharyngeal end. It descends in the lateral wall of the pharynx, inside the constrictor muscles, and blends with the palatopharyngeus muscle. The salpingopharyngeus muscle elevates the pharynx (and larynx) and opens the pharyngeal orifice of the auditory tube during deglutition.

Outer Muscular Layer

Superior Constrictor Muscle

The superior constrictor muscle arises from the posterior border of the lower part of the medial pterygoid plate, the pterygoid hamulus, the pterygomandibular raphe, the posterior part of the mylohyoid line of the mandible, and by some fibers from the side of the tongue. The fibers pass backward in a broad sheet and insert into the median (pharyngeal) raphe. The most superior fibers attach by an aponeurosis to the pharyngeal tubercle of the occipital bone. The palatopharyngeal sphincter, which arises from the anterior and lateral part of the upper surface of the palatine aponeurosis, passes backward, lateral to the levator palati muscle, and blends with the internal surface of the superior constrictor muscle. It produces a rounded ridge on the wall of the pharynx (Passavant's ridge) that is seen upon contraction of the nasopharyngeal sphincter.

Middle Constrictor Muscle

The middle constrictor muscle arises from the posterior edge of the lower part of the stylohyoid ligament and lesser horn of the hyoid bone, and from the upper edge of the greater horn of the hyoid bone. The fibers spread upward from the stylohyoid ligament and lesser horn of the hyoid bone and downward from the greater horn of the hyoid bone, forming a fan-shaped sheet as it passes backward to insert into the whole length of the median raphe. The upper fibers ascend and overlap the superior constrictor muscle, and the middle fibers run horizontally backward. The lower fibers descend to the lower end of the pharynx, passing deep to the inferior constrictor muscle.

Inferior Constrictor Muscle

The inferior constrictor, the thickest of the constrictor muscles, is composed of two parts: thyropharyngeus and cricopharyngeus (**Fig. 46–6**). The thyropharyngeus part of the inferior constrictor arises from the oblique line of the lamina of the thyroid cartilage, from a tendinous band across the cricothyroid muscle, from the lateral surface of the cricoid cartilage at the lower edge of the tendinous band, and from a slip from the inferior horn of the thyroid cartilage. The fibers pass backward and insert into the pharyngeal raphe. The upper fibers ascend obliquely, overlapping the middle constrictor muscle. The cricopharyngeus portion of the inferior constrictor arises from the side of the cricoid cartilage. The fibers run horizontally backward, encircling the junction between the pharynx and esophagus, and insert on the opposite side of the cricoid cartilage. The cricopharyngeus fibers are continuous with the circular fibers of the esophagus and are believed to act as a sphincter, regulating material entering the esophagus. At rest, entry into the esophagus is closed, but with stimulation it becomes open.

Killian's dehiscence is a posterior triangular interval between the upper edge of the cricopharyngeus and lower part of the thryopharyngeus (**Fig. 46–6**). This interval is not a weakness in the pharyngeal wall, but rather a normal characteristic of this region. However, if pressure builds up in the lower part of the pharynx due to incoordination of the pharyngeal peristaltic wave, the deficiency in the constrictor muscles renders Killian's dehiscence a likely place for a diverticulum to form.

Buccopharyngeal Fascia

The buccopharyngeal fascia is a thin fibrous layer on the external surface of the pharynx. It is continuous with the external covering of the constrictor muscles and contains the pharyngeal plexus of nerves and veins. Posteriorly, the buccopharyngeal fascia is attached to the prevertebral fascia; at the sides, it is connected to the styloid process and the carotid sheet. The areolar layer of the buccopharyngeal fascia permits movements of the pharynx. Between the inferior constrictor muscle and the esophagus, only the buccopharyngeal fascia, pharyngobasilar fascia, and underlying mucous membrane make up the pharyngeal wall.

FASCIAL PLANES/SPACES OF THE PHARYNX

Retropharyngeal Space

This potential space consists of loose areolar connective tissue lying between the prevertebral fascia and buccopharyngeal fascia (**Figs. 46–3** and **46–7**). This space may contain the retropharyngeal group of lymph nodes and be important for infection or tumor spread. The space is closed superiorly by the base of the skull and on

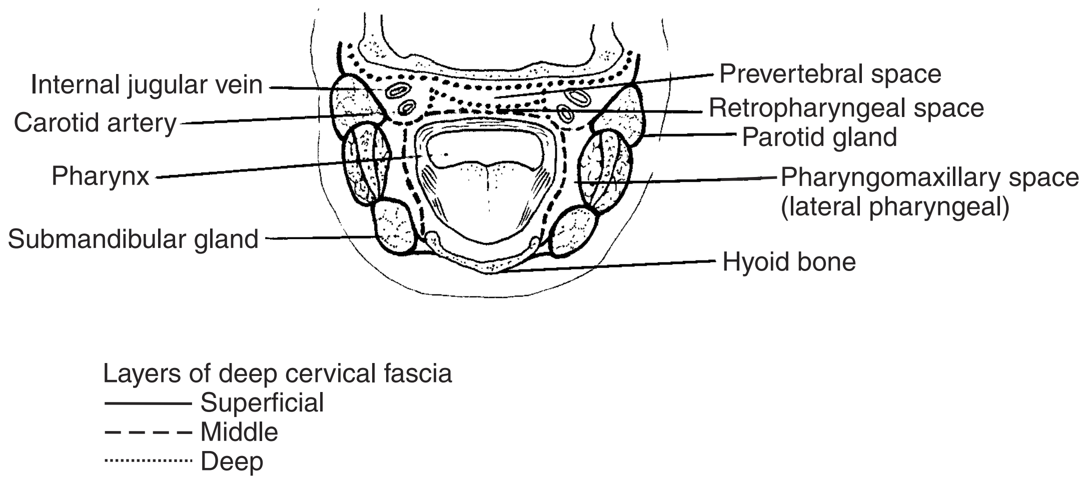


Figure 46–7 *Oblique cross section from the level of the parotid superiorly to the hyoid inferiorly, showing the relationships among the prevertebral, retropharyngeal, and lateral pharyngeal spaces. (Modified from Shumrick KA, Sheft SA. Deep neck infections. In: Paparella MM, Shumrick DA, Gluckman JL, Meyeroff WL, eds. Otolaryngology. Vol 3. Philadelphia: WB Saunders; 1991: Fig. 3, courtesy of the publishers.)*

each side by the carotid sheath. Inferiorly, the retropharyngeal space opens into the superior mediastinum. Infections in the retropharyngeal space may therefore extend into the superior mediastinum. The retropharyngeal space permits movement of the pharynx, larynx, trachea, and esophagus during swallowing and is of considerable surgical importance because of its relationship to these structures and as a potential route of infection spread. A retropharyngeal abcess, formed by pus entering the retropharyngeal space, may produce dysphagia and dysarthria.

Parapharyngeal Space

The parapharyngeal space is an inverted pyramidal-shaped potential space containing areolar tissue and fat and is continuous with the lateral portion of the retropharyngeal space. Lying lateral to each side of the pharynx (**Fig. 46–7**), it extends from the base of the skull to the superior mediastinum. The medial part of the carotid sheath is contained within the posterolateral wall of the parapharyngeal space. On the posteromedial side, a potential communication exists between the parapharyngeal space and the retropharyngeal space (**Fig. 46–7**); anteriorly and inferiorly, the parapharyngeal space communicates with the spaces associated with the floor of the mouth (e.g., submandibular space). Because of these connections and its location, the parapharyngeal space is the space of the neck most often involved with head and neck infections. Furthermore, infections may readily spread from this space to other regions, including the retropharyngeal space and carotid sheath.

The attachment of the pharynx to the base of the skull near the eustachian apparatus is oblique, with the lateral point of attachment on either side being the angle between the spine of the sphenoid and the petrous part of the temporal bone. As a result of this oblique attachment, a lateral recess, the fossa of Rosenmüller, is formed (**Fig. 46–8**). The inner surface of the cartilaginous portion of the eustachian tube can be palpated in the wall of this recess. Therefore, the recess is advantageous in cannulation of the eustachian tube. Lateral to the fossa of Rosenmüller are the internal jugular vein, the internal carotid artery, and cranial nerve (CN) IX, CN X, and CN XI. An aneurysm of the internal carotid may bulge into the lateral recess, resembling a retropharyngeal or peritonsillar abscess.

The parapharyngeal space is divided by the styloid process and the muscles arising from it (stylohyoid, stylopharyngeus, styloglossus) into a prestyloid and poststyloid space. The prestyloid space is bounded medially by the buccopharyngeal fascia and laterally by the medial pterygoid muscle. The glossopharyngeal nerve is situated within the prestyloid space. Anterior to the prestyloid region is the pterygoid process of the parotid gland. Malignant extension of the parotid gland into the poststyloid space is inhibited by the styloid diaphragm. Neurovascular tissues, including the internal carotid artery, internal jugular vein, superior sympathetic ganglion of the cervical chain, and CN VII, CN IX, CN X, CN XI, and CN XII are contained within the poststyloid space. The facial nerve (CN VII) has only a brief course through the poststyloid space as it exits the skull via the stylomastoid foramen.

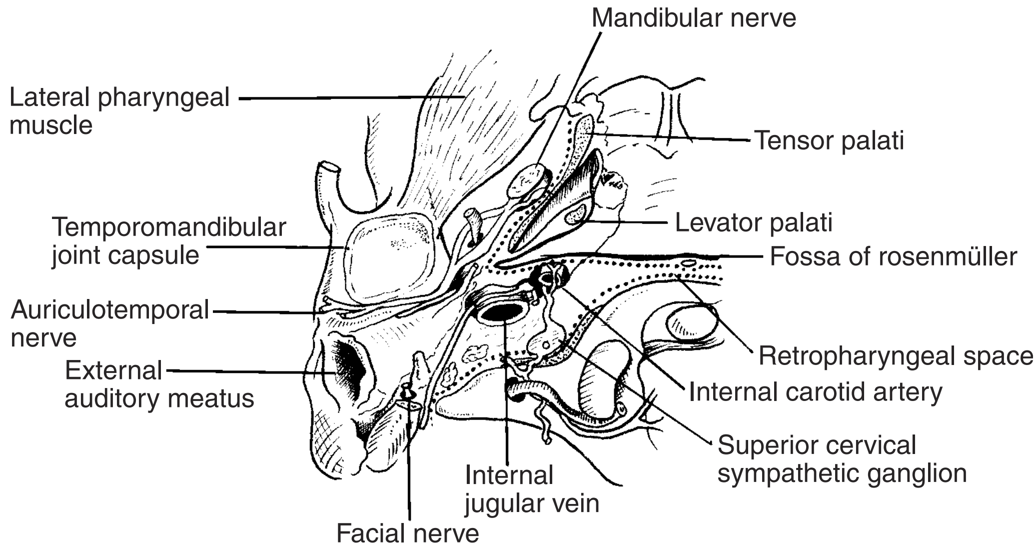


Figure 46–8 *Relationships of the attachment of the pharynx at the base of the skull, demonstrating the contents of the infratemporal fossa (lying anterolaterally), the fossa of Rosenmüller, the internal jugular vein, and the internal carotid artery. (Modified from Davies J, Duckert L. Embryology and anatomy of the head, neck, face, palate, nose, and paranasal sinuses. In: Paparella MM, Shumrick DA, Gluckman JL, Meyerhoff WL, eds. Otolaryngology. Vol 1. Philadelphia: WB Saunders; 1991: Fig. 23, courtesy of the publishers.)*

In the lower part of the neck, where the lateral wall is formed by the sternocleidomastoid and infrahyoid muscles, the parapharyngeal space contains the deep cervical lymph nodes inferiorly bounded by the hyoid. Infection may be carried by lymph vessels that drain into these nodes and may spread to the superior mediastinum. At this level, the poststyloid space also contains the internal jugular vein, internal carotid artery, sympathetic chain, and CN IX to CN XII.

Pteryogomandibular Space

The pteryogomandibular space lies between the medial pterygoid muscle and the mandible. The inferior alveolar nerve and internal maxillary artery pass through the pteryomandibular space. The lingual, glossopharyngeal, and inferior alveolar nerves may become involved with tumors that expand into the parapharyngeal space. These tumors may spread by perineural invasion through the base of the skull.

LYMPHATIC DRAINAGE OF THE PHARYNX

The lymph vessels of the pharynx either directly or indirectly drain into the deep cervical group of lymph nodes. Efferent vessels from this group of lymph nodes form the jugular trunk. On the right side, the jugular trunk either terminates at the junction of the internal jugular and subclavian veins or joins the right lymphatic duct. On the left side, the jugular trunk typically drains into the thoracic duct. However, it may enter the internal jugular or subclavian vein. The deep cervical group of nodes is divided into two groups: a superior group, which is situated next to the internal jugular vein, deep to the sternocleidomastoid muscle; and an inferior group, which is related to the brachial plexus and subclavian vessels and lies partly deep to the sternocleidomastoid muscle. Within the superior group is the jugulodigastric group, which lies in the triangular region bounded by the posterior belly of the digastric muscle, the facial vein, and the internal jugular vein. The jugulo-omohyoid node, lying just above the intermediate tendon of the omohyoid muscle, is within the inferior group of deep cervical lymph nodes.

Nasopharynx

The nasopharynx drains into the retropharyngeal lymph nodes, which lie between the buccopharyngeal fascia and the prevertebral fascia. These nodes atrophy during childhood but may become clinically important in tumor spread. Efferent vessels from these nodes enter the upper deep cervical lymph nodes. Efferent lymphatic flow from the nasopharynx also drains into the jugulodigastric nodes (high level II) and the spinal accessory chain.

Oropharynx

The jugulodigastric nodal group (levels II and III) is the primary drainage site from the oropharynx. Lymphatic vessels from the tonsil pierce the buccopharyngeal fascia and the superior constrictor muscle, then pass between the stylohyoid muscle and internal jugular vein to empty into the jugulodigastric node. The retropharyngeal and pharyngeal nodes drain the pharyngeal portion of the soft palate, the lateral and posterior walls of the oropharynx, and the base of the tongue. These nodes lie in the retropharyngeal and parapharyngeal space in close relation to CN IX to CN XII, the internal jugular vein, and the internal carotid artery. Efferent vessels from these nodes drain to the jugulodigatric and posterior cervical group of lymph nodes.

Laryngopharynx

Primary lymphatic drainage of the laryngopharynx is to the upper cervical lymph nodes. However, the laryngopharynx may also drain to the retropharyngeal nodes and to the paratracheal nodes lying alongside the trachea and esophagus. Efferent vessels from these nodes pass to the deep cervical lymph nodes. The rich lymphatic drainage of the pharynx accounts for the high incidence of metastasis with laryngopharyngeal malignancy.

NERVE SUPPLY OF THE PHARYNX

The pharynx receives its motor, sensory, and autonomic nerve supply through the pharyngeal plexus of nerves. This plexus is situated in the buccopharyngeal fascia surrounding the pharynx and is formed by pharyngeal branches of the glossopharyngeal nerve (CN IX) and vagus nerve (CN X), and by sympathetic fibers from the superior cervical ganglion (**Fig. 46–8**).

The motor fibers in the plexus are derived from the cranial root of the accessory nerve and are transmitted to the pharyngeal muscles via the pharyngeal branch of the vagus nerve, which arises immediately after exiting the skull. With the exception of the stylopharyngeus, all muscles of the pharynx are supplied by the pharyngeal plexus. The stylopharyngeus is supplied by a muscular branch of the glossopharyngeal nerve and is the only muscle supplied by this otherwise sensory nerve. The motor fibers to the constrictor muscles are special visceral (branchial) efferent and arise in the nucleus ambiguus in the medulla. The connections to the nucleus ambiguus are important in reflexes, including coughing, sneezing, swallowing, and gagging. The cricopharyngeus portion of the inferior constrictor muscle receives additional motor innervation from the external laryngeal nerve and parasympathetic fibers from the recurrent laryngeal nerve (**Fig. 46–2**).

The sensory fibers in the pharyngeal plexus are derived from the glossopharyngeal nerve. These fibers supply most of the pharyngeal mucosa. A region of the nasopharynx around the pharyngotympanic (eustachian) tube and lateral pharyngeal recess receives a sensory supply from the pharyngeal branch of the maxillary nerve. In addition, sympathetic and parasympathetic (autonomic) fibers are transmitted by the nerve of the pterygoid canal (vidian nerve). Sensory fibers to the oropharynx are derived from the pharyngeal plexus on the lateral surface of the middle constrictor muscle and by inferior and superior laryngeal branches of the vagus nerve. The superior laryngeal branch of the vagus nerve contributes to the sensory nerve supply of the laryngopharynx.

BLOOD SUPPLY OF THE PHARYNX

The ascending pharyngeal artery, a branch of the external carotid artery, ascends behind the carotid sheath to supply the pharyngeal wall and tonsils (**Fig. 46–2**). The palatine branch of the ascending pharyngeal artery passes over the upper edge of the superior constrictor muscle and supplies the interior of the pharynx and the tonsils. The pharyngotympanic tube is supplied by a small branch of the ascending pharyngeal artery. The tonsillar artery, a branch of the facial artery, passes through the superior constrictor muscle, supplying the inner aspect of the pharynx and entering the inferior pole of the tonsillar bed. The ascending palatine branches of the facial artery, the greater palatine and pterygoid branches of the maxillary artery, and the dorsal lingual branches of the lingual artery provide an additional blood supply to the pharynx.

The pharyngeal veins, which are organized as an internal submucosal and external pharyngeal plexus, drain into the internal jugular and anterior facial veins. The pharyngeal plexus also communicates with the pterygoid plexus. In addition, communicating branches from the plexus anastomose with veins of the dorsum of the tongue, the superior laryngeal veins, and the veins of the esophagus.

STRUCTURAL OVERVIEW OF THE ESOPHAGUS

The esophagus is a muscular tube that connects the pharynx to the stomach. It is ~25 cm long and extends from the lower border of the cricoid cartilage at the level of the sixth cervical vertebra to the cardiac orifice of the stomach. The cervical part of the esophagus lies posterior to the trachea (**Fig. 46–3**) and is attached by loose connective tissue. The right and left recurrent laryngeal nerves ascend in the groove between the

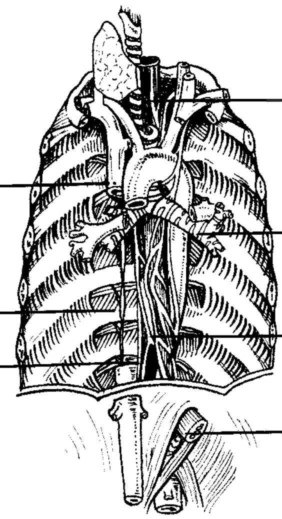


Figure 46–9 *Anterior view of the superior and posterior mediastinum showing the relationship to the esophagus. (Adapted from Beasley P. Anatomy of the pharynx and esophagus. In: Gleeson M, ed. Scott-Brown's Otolaryngology. Bath, England: University of Bath Press; 1997: Fig. 10.30. Reprinted with permission.)*

trachea and esophagus. The cervical esophagus rests on prevertebral fascia that covers the vertebral bodies (C6–C8) and prevertebral muscles. The thoracic duct ascends behind the esophagus along its left side. As it traverses the neck, the esophagus deviates to the left behind the suprasternal notch; therefore, surgical approaches to the cervical esophagus are generally made on the left side. As the esophagus descends through the superior and posterior mediastinum, it rests on the vertebral column, with the distal one third passing just behind the heart. It again curves to the left to cross the thoracic aorta, pierces the diaphragm by passing through the esophageal hiatus, and joins the stomach (**Fig. 46–9**).

In the adult, the diameter of the esophagus is 20 mm, but it may expand to as much as 30 mm to accommodate the passage of a bolus of food. In the newborn, the diameter is ~5 mm. The esophagus, the narrowest part of the alimentary tract, contains three constrictions of anatomical importance: (1) at the upper esophageal sphincter, ~15 cm from the upper incisor teeth; (2) at the anterior compression of the esophagus by the aortic arch and left main bronchus, ~20 to 25 cm from the upper incisors; and (3) at the gastroesophageal junction, ~40 cm from the upper incisor teeth. These narrow regions become of particular concern when dealing with ingestion of foreign matter or caustic agents.

Several areas of inherent weakness are also present in the esophageal wall; for example, below the cricopharyngeus muscle, where the area of the upper posterior esophageal wall is covered only by circular fibers of the esophagus. This region is referred to as the Laimer-Haeckerman triangle (**Fig. 46–2**). Dehiscences that permit the passage of nerves and blood vessels also serve as areas of potential weakness. These include the lateral dehiscence between the cricopharyngeus muscle and circular fibers of the esophagus (Killian-Jamieson area), through which the recurrent laryngeal nerve passes (**Fig. 46–2**).

THE ESOPHAGEAL WALL

The esophageal wall is composed of four layers: mucosa, submucosa, muscle layer, and outer fibrous layer.

MUCOSA

A tough, nonkeratinized stratified squamous epithelium lines the esophagus. This epithelium, which is continuous with that of the pharynx, has a basement membrane overlying a loose areolar connective tissue lamina propria rich in collagen and elastic fibers, blood vessels, nerve fibers, and lymphoid nodules. As shown by a normal barium swallow, the mucous membrane consists of seven to 10 longitudinal folds that disappear during dilatation. The mucosa has a thin smooth muscle layer (muscularis mucosae) that thickens toward the lower end of the esophagus. At the level of the cricoid cartilage, the muscularis mucosae is continuous with the elastic layer of the pharynx. A transition from nonkeratinized, squamous epithelium to columnar epithelium takes place at the junction of the esophagus with the stomach.

SUBMUCOSA

The submucosa is a thick layer of dense collagenous connective tissue. Like the mucosa, it is thrown into longitudinal folds when the esophagus is at rest. The submucosa contains blood vessels, Meissner's nerve plexus of postganglionic parasympathetic fibers, and esophageal glands that lubricate the passage of food by mucous secretion.

Muscle Layer (Tunica Muscularis)

The muscles of the esophagus are arranged into an outer longitudinal and inner circular layer (**Fig. 46–2**). The longitudinal fibers form a complete covering for almost the entire esophagus, except for a region 3 to 4 cm below the cricoid cartilage, where the fibers diverge to form two longitudinal fasciculae that pass upward and forward to attach to the lamina of the cricoid cartilage by a tendon (**Fig. 46–2**).

The circular layer is continuous with the cricopharyngeus part of the inferior constrictor muscle. Anteriorly, the fibers of the circular layer insert into the tendon of the longitudinal muscle.

Inferiorily, these fibers are continuous with the muscle fibers of the stomach. At the distal end of the esophagus, the circular fibers form part of a "physiological" sphincter.

The myenteric plexus of Auerbach is situated between the longitudinal and circular layers. In the upper third of the esophagus, the fibers of both layers are striated; in the middle third, the fibers are mixed striated and smooth muscle; in the lower third, the fibers are nonstriated, which has implications for connective tissue diseases such as scleroderma.

Outer Fibrous Layer

The outer fibrous layer is composed of an external adventitia of irregular, dense connective tissue containing blood vessels, nerve fibers, and a complex network of elastin fibers. The elastin fibers are continuous with fibers within the inner layers of the esophagus, permitting expansion of the esophagus during deglutition. The presence of the adventitia facilitates mobilization of the esophagus from above or below without opening the thoracic cavity, as, for example, in a translateral pharyngolaryngoesophagectomy.

BLOOD SUPPLY OF THE ESOPHAGUS

The cervical part of the esophagus obtains its blood supply from the inferior thyroid arteries that originate from the thyrocervical trunks of the subclavian artery. The thoracic part is supplied either directly from the thoracic aorta or indirectly by branches of the bronchial or upper posterior intercostal arteries. The abdominal part of the esophagus derives its blood supply from esophageal branches of the left gastric artery, which arise from the celiac trunk, and the left inferior phrenic artery, which arises from the abdominal aorta.

A venous plexus is present along the exterior of the esophagus and drains in a manner that parallels the blood supply. The cervical part of the esophagus drains into the inferior thyroid veins, and the thoracic part drains into the azygos/hemiazygos system. The left gastric vein drains the abdominal part of the esophagus. Because the left gastric vein is a tributary of the portal system, the abdominal region of the esophagus is an important site for portal-systemic anastomoses.

LYMPHATIC DRAINAGE OF THE ESOPHAGUS

The cervical part of the esophagus drains into the deep cervical nodes and the paratracheal nodes. The thoracic part drains into tracheobronchial nodes and posterior mediastinal nodes. The abdominal part of the esophagus drains into left gastric nodes and celiac nodes. Some efferent vessels from this lower region of the esophagus may drain directly into the thoracic duct.

NERVE SUPPLY OF THE ESOPHAGUS

Branches of the recurrent laryngeal nerve supply the striated muscle of the cervical part of the esophagus. The cell bodies for these nerve fibers are found in the nucleus ambiguus. The main motor supply to the nonstriated muscle is parasympathetic. The cell bodies for these fibers are in the dorsal nucleus of the vagus. The nerve fibers reach the esophagus through esophageal branches of the vagus nerve and through its recurrent laryngeal branches (**Fig. 46–2**). Located between the outer longitudinal and inner circular muscle layers are the ganglia of Meissner's plexus and Auerbach's plexus. After synapsing in these ganglia, postganglionic fibers innervate the muscle fibers.

The lateral gray column of the spinal cord (thoracic segments 2–6) contains the cell bodies of the preganglionic sympathetic motor fibers. These fibers reach the sympathetic trunk through white rami communicantes, synapse in the cervical ganglia, then reach the esophagus as postganglionic fibers by traversing the cardiac plexus. In the cervical esophagus, a plexus forms around the inferior thyroid artery; in the thoracic esophagus, a plexus forms around the branches of the descending thoracic aorta, bronchial arteries, and upper intercostal arteries. Plexuses that form around the left gastric and inferior phrenic arteries supply the abdominal esophagus.

Afferent fibers from the esophagus travel with branches of the vagus nerve. Their cell bodies lie in the inferior vagal ganglion. Pain sensation may be conveyed by some of the afferent fibers that travel with the sympathetic nerves.

Suggested Readings

Ballenger JJ. Anatomy and physiology of the oral cavity and pharynx. In: Ballenger JJ, Snow JJ Jr, eds. Otorhinolaryngology: Head and Neck Surgery. Baltimore: Williams & Wilkins; 1996:220–227

Beasley P. Anatomy of the pharynx and esophagus. In: Gleeson M, ed. Scott-Brown's Otolaryngology. Bath, England: University of Bath Press; 1997:10/1–10/40

Bosma JF, Donner MW. Physiology of the pharynx. In: Paparella MM, Shumrick DA, Gluckman JL, Meyerhoff WL, eds. Otolaryngology. Vol 1. Philadelphia: WB Saunders; 1991: 371–389

Davies J, Duckert L. Embryology and anatomy of the head, neck, face, palate, nose, and paranasal sinuses. In: Paparella MM, Shumrick DA, Gluckman JL, Meyeroff WL, eds. Otolaryngology. Vol 1. Philadelphia: WB Saunders; 1991:59–106

Ger R, Abrahams P. Upper respiratory and alimentary passages. In: Ger R, Abrahams P, eds. Essentials of Clinical Anatomy. New York: Churchill Livingstone; 1989:255–261

Harrison DFN. Tumors of the hypopharynx. In: Paparella MM, Shumrick DA, Gluckman JL, Meyeroff WL, eds. Otolaryngology. Vol 3. Philadelphia: WB Saunders; 1991:2199–2214

Seiden AM. Esophageal disorders. In: Paparella MM, Shumrick DA, Gluckman JL, Meyeroff WL, eds. Otolaryngology. Vol 3. Philadelphia: WB Saunders; 1991:2439–2482

Shumrick KA, Sheft SA. Deep neck infections. In: Paparella MM, Shumrick DA, Gluckman JL, Meyeroff WL, eds. Otolaryngology. Vol 3. Philadelphia: WB Saunders; 1991:2545–2563

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. The space of the neck most frequently involved with head and neck infections is the

   A. Retropharyngeal space
   B. Parapharyngeal space
   C. Pterygomandibular space
   D. Sinus of Morgagni

2. Areas of inherent or potential weakness in the esophageal wall are present

   A. Below the level of the cricopharyngeus muscle
   B. In the Laimer-Haeckerman triangle
   C. In the Killian-Jamieson area
   D. A and B only
   E. A, B, and C

3. In patients with esophageal reflux, the following may develop:

   A. Hypertrophy of the fibers of the cricopharyngeal sphincter
   B. A rise in intraluminal pressure in the hypopharynx
   C. Outpouching of the mucosa in Killian's dehiscence
   D. All of the above

Chapter 47
THE BIOLOGY OF SWALLOWING

SOLY BAREDES AND KRISTINE MOISER

REVIEW OF RELEVANT ANATOMY AND PHYSIOLOGY

GENERAL ANATOMICAL RELATIONSHIPS OF THE UPPER AERODIGESTIVE TRACT

The act of deglutition involves transporting a liquid or a solid bolus from the anterior oral cavity through the lower esophageal sphincter (LES), where the esophagus empties into the cardiac orifice of the stomach. Thus the bolus must traverse an inverted L-shaped passage, and simultaneously must be prevented from penetrating the nasopharyngeal and laryngeal airways. Successful transport, then, depends on the coordinated actions of structures from four different components: the oral cavity, pharynx, larynx, and esophagus.

The oral cavity is bounded anteriorly by the lips and includes the maxilla and mandible, dentition, hard palate, anterior two thirds of the tongue, floor of the mouth and salivary glands with supplying neurovascular bundles, and overlying mucosa of stratified squamous epithelium. Laterally, the oral cavity is bounded by the buccinator muscle. Posteriorly, the oral cavity joins the oropharynx at the junction of the hard and soft palate.

The pharynx is subdivided into the nasopharynx, oropharynx, and hypopharynx. The nasopharynx is bounded superiorly by the sphenoid sinus and superior clivus, anteriorly by the posterior nasal margin, and posteriorly by the lower clivus and upper cervical spine. The lateral extent of the nasopharynx is bounded by the pterygoid plates more anteriorly, and posteriorly the lateral walls are formed by the muscles of the superior constrictor and fascial planes. Inferiorly, the nasopharynx extends to the level of the soft palate. The oropharynx extends from the soft palate to the epiglottis, with the anterior demarcation at the circumvallate papillae of the tongue (posterior one third or base of the tongue) and the posterior extent at the pharyngeal wall composed of the constrictor muscles overlying the upper cervical spine. The oropharynx includes the soft palate and palatine tonsils and is laterally bounded by the constrictor muscles. The hypopharynx extends from the laryngeal surface of the epiglottis superiorly to the inferior margin of the cricopharyngeus at the junction of the esophagus inferiorly. Anteriorly, the hypopharynx is bounded by the pyriform sinuses, the postcricoid area, and the posterior wall of the larynx. Like the oropharynx, the hypopharynx is bounded laterally by the constrictor muscles and posteriorly by the posterior pharyngeal wall.

The pharyngoesophageal junction is characterized more functionally than structurally with the upper esophageal sphincter (UES). The UES is composed of the cricopharyngeus muscle, which consists of ~7 mm wide band of horizontal and oblique muscle fibers from the inferior constrictor and the upper esophagus. The esophagus is a long, semiflexible muscular tube, ranging in length from ~20 to 24 cm in adults. Inferiorly, the esophagus terminates at the LES, which, like the UES, is a functionally distinct but structurally indistinct zone of ~2 to 4 cm in the esophagogastric region.

Musculoskeletal Components

The complex anatomical relationships of the upper aerodigestive tract and their role in swallowing can perhaps best be understood in the context of the two primary physiological events in swallowing: transport of the bolus from the oral cavity through the esophagus, and protection of the nasopharyngeal and laryngeal airways. From this standpoint the musculoskeletal components involved with swallowing can be categorized according to their respective functions with regard to bolus transport and airway protection. **Table 47–1** illustrates the various muscles of the oral cavity, pharynx, larynx, and esophagus and their functions in swallowing.

Peripheral Nervous System Components

The peripheral nervous innervation of the oral cavity, pharynx, and esophagus is provided through cranial nerve (CN) V, CN VII, CN IX, CN X, and CN XII, with central afferent connections to the brainstem nuclei, the nucleus of the tractus solitarius (NTS), and the nucleus ambiguus. **Table 47–2** illustrates the motor efferent and sensory afferent pathways supplying the oral cavity, pharynx, larynx, and esophagus.

Central Nervous System Components

The control of complex movements such as swallowing requires involvement of several different components of the central nervous system at the cortical, subcortical, and brainstem level. In general, the control and regulation of movement at the central level involves sensory input areas, motor output areas, sensorimotor integration areas, and association areas. These relationships are diagrammed in **Figs. 47–1** and **47–2**.

The sensory input areas of the cortex include the primary and secondary sensory cortices (SI and SII). Motor output from the cortex is effected from the primary motor cortex, M1. In each case, sensory input and motor output are modulated through sensorimotor integration nuclei in the thalamus. In addition, the basal ganglia play a role in modulating motor output. A variety of other cortical connections to the primary sensory and motor cortices [e.g., posterior parietal cortex, sensory association cortex, and supplementary motor area (SMA)] provide the

Table 47–1 Muscular components and their function in swallowing

Function	Oral Cavity and Oropharynx	Pharynx	Larynx	Esophagus
Bolus procurement and manipulation	Muscles of facial expression Muscles of mastication Intrinsic and extrinsic tongue muscles			
Bolus transport	Intrinsic and extrinsic tongue muscles Suprahyoid muscle	Pharyngeal constrictors, Cricopharyngeus		Outer longitudinal Inner circular
Airway protection: Seal nasopharynx Seal pharyngeal inlet Laryngeal airway	Extrinsic tongue muscles Suprahyoid muscle Hyoglossus	Levator and tensor veli palatine Palatoglossus Palatopharyngeus Pharyngeal muscle Aryepiglottis muscle	Intrinsic laryngeal muscles Infrahyoid strap muscle	

TABLE 47–2 PERIPHERAL NERVOUS SYSTEM INNERVATION OF THE OROPHARYNX, PHARYNX, LARYNX, AND ESOPHAGUS

Motor Efferent Pathways			Sensory Afferent Pathways	
Region	**Muscular Components**	**Cranial Nerve (CN)**	**Sensory Receptors**	**Cranial Nerve (CN)**
Oral cavity and oropharynx	Muscles of facial expression Muscles of mastication Intrinsic and extrinsic tongue muscles Suprahyoid muscles Pharyngeal muscles	CN VII CN V, 3rd division CN XII CN V, 3rd division, CN XII CN V, 3rd division, C1 via CN XII, CN VII, CN IX	Somatic sensory: anterior ⅔ tongue Somatic sensory: posterior ⅓ tongue Proprioception Mechanoreceptors	CN V to main sensory or spinal nucleus CN V CN IX to spinal nucleus of CN V CN V to mesencephalic nucleus CN V CN IX to NTS and spinal nucleus of CN V
Pharynx	Palatal muscles Pharyngeal muscles (constrictors) Suprahyoid muscles	Pharyngeal plexus CN X, CN V Pharyngeal plexus CN X CN V, 3rd division, CN XII	Pharyngeal wall receptors	CN X to NTS
Larynx	Cricothryoid muscles All other intrinsic muscles Infrahyoid muscles Thyrohyoid	Ext. br. sup. Laryngeal CN X Recurrent laryngeal CN X C1–3 via ansa cervicalis C1 via CN XII	Laryngeal receptors: Epiglottis to vocal folds Below vocal folds	Intrinsic laryngeal, CN X to NTS Recurrent laryngeal, CN X to NTS
Esophagus	Striated muscle zone Smooth muscle zone	CN X from nucleus ambiguous Dorsal motor nucleus CN X	Esophageal wall receptors	CN X to NTS

NTS, nucleus of the tractus solitarius.

sensory context or the planning information to coordinate the appropriate motor output with the sensory input.

In the brainstem, the nuclei of CN V, CN VII, CN IX, CN X, and CN XII are distributed in a rostral to caudal dimension throughout the pons and medulla. In the medulla, the hypoglossal nucleus sits medial to the dorsal nucleus of the vagus, which in turn sits ventral medial to the NTS. Ventral to these three nuclei is the nucleus ambiguus coursing from the mid to lower medulla.

BASIC MECHANISMS IN SWALLOWING

PHASES OF SWALLOWING

Bolus Preparation: Oral Preparatory Phase

Bolus preparation begins with the introduction of the bolus into the oral cavity. For a solid bolus, mastication must be invoked to prepare the bolus properly for swallowing. Mastication requires physically shearing, macerating, and grinding the solid mass through the actions of the dentition, while at the same time, mixing the solid mass with saliva. Both mastication and proper mixing of the mass with saliva require manipulation by the tongue and muscles of mastication (masseter, temporalis, medial and lateral pytergoids) and facial muscles (orbicularis oris, buccinator).

Manipulation of the bolus, whether liquid or solid, additionally serves to stimulate receptors distributed throughout the oral cavity that are believed to play a role in initiating the swallow.

Oral Phase

Following preparation of the bolus, the bolus must be transported through the oral cavity to the pharynx. Transport of the bolus through the oral cavity occurs primarily through contraction of the intrinsic tongue muscles

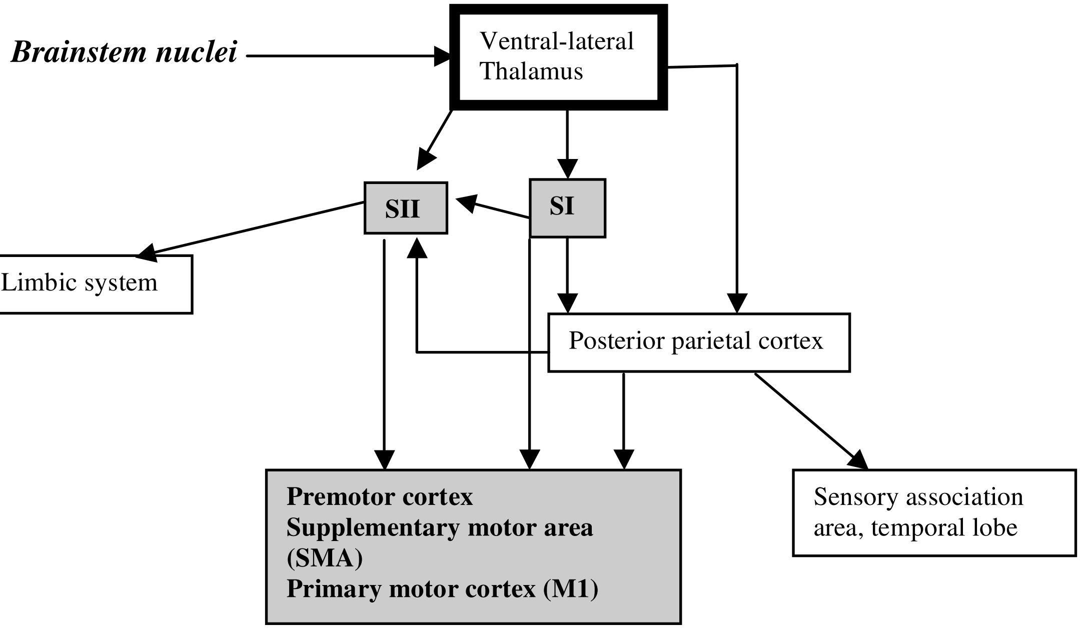


Figure 47–1 *Generalized schematic diagram of sensory input and output pathways to and from the primary (SI) and secondary (SII) sensory cortices. Input from brainstem nuclei is relayed through the thalamic nuclei. Intra- and interhemispheric connections include the sensory association areas of the posterior parietal cortex and temporal lobe, as well as the limbic system [insular cortex, amygdala, and hypothalamus (not shown)]. Output pathways are directed to the premotor areas, supplementary motor area (SMA), and the primary motor cortex (M1).*

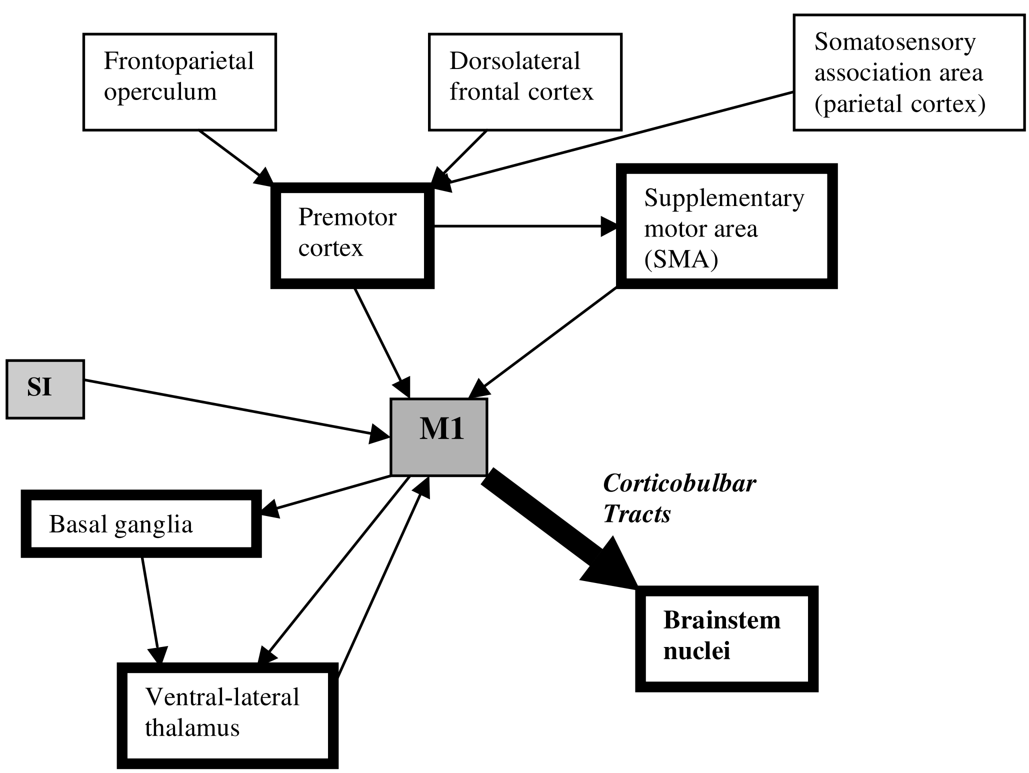


Figure 47–2 *Generalized schematic diagram of the motor input and output pathways to and from the primary motor cortex (M1). Input to M1 comes from motor planning sites (premotor cortex and SMA), the primary sensory cortex (SI), and other sensory association or "perceptual" association areas (somatosensory association area, dorsolateral prefrontal cortex). Output from M1 is modulated via loops through the basal ganglia and integrated with relays through the thalamic nuclei. The major output tract from M1 for swallowing behaviors is through the corticobulbar tracts to brainstem nuclei.*

(transverse, longitudinal, vertical) and extrinsic tongue muscles (hyoglossus, genioglossus, styloglossus). The genioglossus acts to protrude and depress the body of the tongue. The hyoglossus depresses the tongue, whereas the styloglossus elevates and retrudes the tongue. Bolus transport, then, relies on the complex neuromuscular coordination of the intrinsic and extrinsic tongue muscles.

Generally, the oral phase is described as beginning with the tongue tip occluded against the lingual aspects of the anterior teeth. The bolus is then compressed against the hard palate as it is transported through the oral cavity. While the bolus is contained between the dorsum of the tongue and the hard palate, a bulging occurs in the posterior pharyngeal wall (superior constrictor) known as Passavant's ridge. At the same time, the soft palate is opposed against the raised base of the tongue, and the faucial dimension is reduced, both by action of the palatoglossus. This prevents the bolus from prematurely entering the pharynx. Traditionally, movements of the bolus by actions of the tongue are described by a midline depression forming in the tongue and a stripping action of the tongue against the palate, whereby the bolus is then funneled or thrust to the pharyngeal inlet.

Pharyngeal Phase

The pharyngeal phase begins with the presentation of the bolus at the faucial arches. At this point, several events occur simultaneously: (1) elevation of the soft palate to seal off the nasopharynx, (2) glottal adduction, (3) elevation of the hyoid/laryngeal complex, and (4) constriction of the pharyngeal walls to drive the bolus into the esophagus. Specifically, contraction of the levator veli palatini elevates the soft palate, while at the same time the tensor veli palatini tenses and flattens out the palate. Both actions of the palatal muscles serve to seal off the nasopharynx from the oropharynx.

Glottal adduction is achieved through contraction of the intrinsic laryngeal muscles, primarily the vocalis muscle, the lateral cricoarytenoid (LCA), and the cricothyroid in concert with inhibition of the posterior cricoarytenoid. The synergistic activity of these muscles provides protection of the glottis at two levels. Adduction of the paired vocalis muscles seals off the glottis at the level of the true cords. Contraction of the LCAs acts on the cricothyroid joint along with the criocothyroid muscle, which elongates the chink and draws the two cords together, thus sealing off the glottis at the level of the laryngeal vestibule. In addition, the arytenoid cartilages tip anteriorly and superiorly to pull in the aryepiglottic folds, and at the same time, the base of tongue moves posteriorly to obstruct the laryngeal vestibule.

Elevation of the hyoid/laryngeal complex is likewise important for protection of the glottal airway. Elevation of the hyoid/laryngeal complex serves two purposes: (1) to clear the larynx out of the path of the bolus and (2) to provide a mechanism for passive opening of the upper esophageal sphincter. The suprahyoid muscles are responsible for elevation and stabilization of the hyoid and larynx. Displacement of the hyoid and larynx anteriorly and superiorly also serves to facilitate epiglottic tilt. During the anterior and superior excursion of the hyoid and larynx, the epiglottis is tilted posteriorly and inferiorly over the laryngeal vestibule; this covers the vestibule and acts to project the bolus away from the larynx and into the pharyngeal lumen. Epiglottic tilt, however, does not completely cover the laryngeal aditus, and food particles are able to penetrate this opening. Return of the hyoid and larynx to the resting position occurs following clearance of the bolus from the pharynx, which is facilitated by the middle and inferior constrictors and by contraction of the infrahyoid strap muscles.

Movement of the bolus through the pharynx is initially accomplished by contraction of the semiconcentric superior constrictor. The palatopharyngeus, stylopharyngeus, and salpingogpharyngeus dilate, raise up, or pull the pharynx laterally, respectively, which aids the constrictor in moving the bolus inferiorly. The styloglossus, palatoglossus, stylohyoid, and posterior belly of the digastric are active at the same time as the superior constrictor and seal off the pharyngeal inlet to prevent regurgitation into the oral cavity. In the lower pharynx, contraction of the middle and inferior constrictors continues the peristaltic wave of contractions initiated by the superior constrictor.

A small (~7 mm) band of fibers from the inferior portion of the inferior constrictor and the superior portion of the esophageal constrictor constitutes the cricopharyngeus muscle, which attaches to the cricoid cartilage. The cricopharyngeus muscle acts as a muscular valve [the UES, or the pharyngoesophageal (PE) segment] to direct passage of the bolus from the pharynx into the esophagus. In the absence of stimulus the UES remains closed under tonic contraction of the cricopharyngeus. The anterior and superior displacement of the larynx, via attachments of the cricoid cartilage, results in opening of the cricopharyngeus due to stretching of the muscle fibers and in widening the pharyngeal lumen due to the displacement of the cricoid cartilage. Inhibition of tonic contraction of the muscular sphincter occurs prior to opening of the cricopharyngeus (usually 200–300 msec after swallow initiation) and is believed to facilitate the opening.

To ensure appropriate transport of the bolus while preventing penetration of the airways, swallowing and respiration must be coordinated in the adult human. Thus swallowing occurs most commonly during expiration and lengthens the expiration phase. When a swallow occurs during inspiration, the inspiration is terminated, thereby shortening the inspiratory phase. The timing of the swallow during expiration or inspiration is believed to facilitate the intraluminal pharyngeal pressure differences that assist in propelling the bolus. A swallow at the end of inspiration or at the beginning of expiration, for example, is associated with a significant negative pressure in the pharynx that is thought to pull the bolus through the pharynx to the UES.

Esophageal Phase

The esophagus is structurally differentiated into two portions: the proximal one third of the esophagus, which is composed predominantly of striated muscle, and the distal one half to two thirds that is composed of smooth muscle. Passage of the bolus through the UES into the esophagus begins the esophageal phase. Transport through the esophagus is achieved by peristaltic waves of the striated and smooth muscular portions in a rostral to caudal direction. Peristalsis occurs as primary and secondary waves. Primary peristalsis is characterized by an initial rapid phase of inhibition, followed by a more prolonged wave of contraction that traverses the length of the esophagus. During primary peristalsis, the LES remains open, generally ~6 to 8 seconds. The LES, like the UES, is an anatomically undefined but manometrically functional muscular valve consisting of a high-pressure zone of ~2 to 4 cm in the esophagogastric region. Secondary peristalsis occurs following primary peristalsis and is believed to occur in response to local stimulation (e.g., distention) and consists of a rostral-caudal propagating wave of contractions.

Neural Control of Swallowing

At present, very little is known about the control of swallowing at the cortical level. Nevertheless, both human and animal studies have demonstrated that the primary motor cortex is involved with the control of swallowing behaviors. More recent studies have shown that multiple cortical areas are activated during swallowing, including the primary sensory and motor cortices, the supplementary motor area, and the insular cortex. The SMA is believed to play a role in planning of sequential movements as occurs with the transition of neuromuscular events from the oral cavity through the esophagus. The insular cortex is involved with a variety of sensorimotor functions, including sensorimotor integration between M1 and thalamic nuclei, and plays a role in processing of "routine" or "overlearned" motor tasks, which includes swallowing. In addition, these studies have demonstrated activity in the thalamus and basal ganglia during swallowing tasks, reflecting the role of these subcortical sites in sensorimotor integration.

Ascending and descending information to and from the cortex and the brainstem is relayed via the corticobulbar fibers through the internal capsule, which also has been demonstrated during swallowing tasks. In the brainstem, the critical input and output centers for swallowing reside in the nucleus of the NTS and the nucleus ambiguus, respectively. The NTS is organized somatotopically such that representations for the mouth to the esophagus are arranged in a rostral to caudal dimension. Sensory input from cranial nerves synapses on the NTS. The NTS projects to a central pattern generator associated with swallowing in the adjacent reticular formation. Through this projection, the NTS is believed to generate the reflexive patterned response to sensory input that is then output to motoneurons of CN V, CN VII, CN XII, and the nucleus ambiguus. Fibers from the nucleus ambiguus are output through CN IX (to the stylopharyngeus), via fibers of CN XI to CN X, or directly via CN X, to the palatal muscles (except the tensor, CN V) and to the muscles of the pharynx and larynx (see **Table 47–2**). Motor output to the esophagus is divided between the dorsal nucleus of the vagus, which supplies the lower smooth muscle portion and the LES, and the nucleus ambiguus supplies the rostral striated portion of the esophagus.

CONSIDERATIONS IN CLINICAL EVALUATION OF SWALLOWING: APPLICATIONS OF BIOLOGICAL PRINCIPLES

Clinical evaluation of swallowing function in patients presenting with a history or symptoms of dysphagia begins with a careful history of the patient's feeding habits (diet, feeding method, duration of meals, weight loss, or coughing or choking during meals); history of persistent cough, hoarseness, sore throat, heartburn or regurgitation; relevant medical history (history of surgery or radiation treatment in the head and neck, neurological/neuromuscular disorders, previous episodes of aspiration pneumonia, history of gastroesophageal reflux); medications (e.g., antihistamines, antipsychotics, antidepressants, and diuretics); and other social or behavioral factors such as alcohol or tobacco use.

Oral Function

Clinical evaluation of oral function in swallowing primarily focuses on assessment of the ability to perform maneuvers utilized for bolus procurement and transport. Patients should be evaluated for mental status, sensory disturbances, oromotor function (facial muscles, muscles of mastication, tongue muscles), voice quality, dental condition, and alterations in mucosal membranes. Abnormalities of the oral phase include drooling, repetitive tongue movements with failure to propel bolus, inability to transfer bolus to oropharynx, nasal regurgitation, and premature spillage into the oropharynx.

Pharyngeal Function

Assessment of the pharyngeal phase focuses on transport and clearance of the bolus, as well as airway protection, and includes evaluation of palatopharyngeal closure, pharyngeal contraction, hyoid/laryngeal elevation (including evaluation of the intrinsic and extrinsic laryngeal muscles), and function of the cricopharyngeus. On clinical examination, evaluation of palatopharyngeal closure should be observed during rest, during phonation, and with stimulation of the gag reflex. Pharyngeal contraction should likewise be evaluated during stimulation of the gag reflex. Hyoid/laryngeal elevation should be clinically assessed initially by palpation. Function of the intrinsic laryngeal musculature, initiation of the pharyngeal phase, and control of the bolus can be assessed via flexible fiberoptic examination. Cricopharyngeal function is best assessed with dynamic radiographic studies, although intraluminal manometry also provides useful information. Abnormalities of the pharyngeal phase would present as nasal regurgitation (due to palatal incompetence, limitation of pharyngeal constrictors), retention of the bolus in the valleculae or pyriform sinuses (due to poor or absent pharyngeal contraction, asymmetric epiglottic tilt, or cricopharyngeal abnormality), and penetration of the laryngeal airway or aspiration (due to poor or absent laryngeal elevation, failure of epiglottic tilt, failure of vocal cord adduction, or poor sensation). Laryngeal sensation can be assessed by stimulation of the larynx directly by fiberoptic endoscopy.

Esophageal Function

Esophageal phase function is centered on assessment of peristalsis through the esophagus as well as function of the LES, and as such, esophageal examination is primarily performed radiographically, with endoscopy and manometry performed where indicated. Morphological analysis of the esophagus is performed to determine the presence of surface abnormalities due to inflammatory lesions, gastroesophageal reflux, esophageal rings or strictures, or tumors. Abnormalities of esophageal motility present as defective or absent primary peristalsis, increased contractions or spasm, and abnormalities of either the cricopharyngeus or the LES.

Suggested Readings

Bass NH. The neurology of swallowing. In: Groher ME, ed. Dysphagia: Diagnosis and Management. Bosston: Butterworth-Heinemann; 1997:7–34

Cunningham ET, Donner MW, Jones B, Point S. Anatomical and physiological overview. In: Jones B, Donner MW, eds. Normal and Abnormal Swallowing Imaging in Diagnosis and Therapy. New York: Springer-Verlag; 1991:7–26

Krepsi YP, Blitzer A, eds. Aspiration and swallowing disorders. In: The Otolaryngological Clinics of North America. Philadelphia: WB Saunders; 1988:613–720

Mendelsohn M. New concepts in dysphagia management. J Otolaryngol 1993;(Suppl.1):5–24

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. The pharyngoesophageal segment
   A. Is defined by the width of the cricopharyngeus muscle
   B. Actively widens during the pharyngeal phase of swallowing
   C. Is in tonic contraction throughout the pharyngeal phase of swallowing
   D. Is widened by the anterior and superior displacement of the larynx during the pharyngeal phase of swallowing

2. Which of the following does not contribute to protection of the airway during swallowing?
   A. Elevation of the hyoid/laryngeal complex
   B. Glottal adduction
   C. Contraction of the cricopharyngeus muscle
   D. Coordination of respiration

3. Bolus progression during the pharyngeal phase
   A. Is primarily a function of gravity
   B. Is achieved by peristalsis of the constrictor muscles and pharyngeal pressure differences
   C. Is primarily a function of the pressure generated by the action of the base of the tongue
   D. Is accomplished by the pharyngeal pressure changes caused by respiration

4. The neural control of swallowing
   A. Does not involve the primary motor cortex
   B. Involves activation of the primary motor and sensory cortices
   C. Is limited to the brainstem
   D. Is strictly reflexive

Chapter 48
LARYNGEAL PATHOLOGY

MAJORIE BRANDWEIN-GENSLER

The upper aerodigestive tract plays host to a vast array of pathological processes of infectious, reactive, and neoplastic etiologies. It would be impossible to cover the range of processes encountered in this compact, crucial region and do justice to their diagnostic criteria and inherent diagnostic dilemmas within these pages. This chapter covers some of the more common tumor and tumor-like processes otolaryngologists may encounter.

THE LARYNX

BENIGN LARYNGEAL TUMORS

Vocal Cord Nodules/Polyps/Contact Ulcers

Vocal cord nodules (laryngeal nodules, or singer's, preacher's, or screamer's nodes) are stromal reactions within Reinke's space, usually trauma related, caused by vocal abuse, although smoking may also play some role. Vocal polyps are usually unilateral, with a male predominance. In contrast, vocal nodules are usually bilateral and symmetrical, with a female predominance. Vocal cord polypoid degeneration, or Reinke's edema, causes bilateral diffuse vocal polyps in the middle-aged to elderly population; it is unrelated to vocal abuse, but its associated with smoking. Nodules and polyps usually occur between the ages of 20 and 60, but nodules can be seen in children. The most frequent symptom is hoarseness. The involved site is the vibratory surface of the true vocal cord, usually the junction of the anterior and middle third, which is the point of maximum vibratory impact. They may appear gray or white, translucent, sessile, or polypoid, and usually measure a few millimeters in diameter.

Contact ulcers (contact granulomas) are distinguished from vocal cord nodules/polyps in that they occur on the vocal process of the arytenoids, as a result of forceful apposition during vocalization. They may be unilateral or bilateral. There is a male predominance, with a predisposition for lawyers, salespeople, managers, and preachers ("preacher's nodules") who must produce a "deep," low-frequency, forceful voice.

Pathology Microscopically, vocal cord nodules and polyps are characterized by the finding of a sparsely cellular stromal change, either myxoid or edematous, fibrous, vascular, and fibrinous or hyaline, underlying the stratified squamous epithelium. Dilated vascular spaces or foci of hemorrhage may be present. Inflammatory cell infiltrates are infrequent, and glandular elements are absent. The squamous epithelium may be normal, atrophic, or keratotic or at times dysplastic. Ulceration is infrequent. The presence of atypical stromal cells has been observed sporadically in vocal cord polyps.

Prognosis Treatment for vocal nodules is voice therapy to eliminate the source of voice abuse. Vocal polyps are excised. These lesions have a benign clinical course but may persist if the etiologic factors remain.

Cysts: Ductal, Oncocytic, Saccular, and Laryngoceles

Laryngeal cysts can be classified as (1) ductal cysts, which are squamous, oncocytic, or "tonsillar" mucin-filled cysts arising from minor salivary ducts obstruction; (2) saccular cysts, arising from obstruction of the saccule; and (3) laryngoceles, which are air-filled pulsion diverticulum of the saccule. In a review of two decades of laryngeal cysts (190 cases) from the Mayo Clinic, ductal cysts were most commonly encountered (75%), usually in the vocal cords and lingual epiglottis.

Ductal Cysts

Laryngeal cysts may result from the blockage of a minor salivary gland duct. In this case, the cyst lining is actually the dilated ductal epithelium. It may be squamous, oncocytic, or squamous with surrounding lymphoid stroma (referred to as "tonsillar" cysts). Squamous cysts and oncocytic cysts have a predisposition for the ventricular bands, ventricle, aryepiglottic folds, and epiglottis. Tonsillar cysts have a predisposition for the valecullae, an area with tonsillar remnants. Simple conservative excision is curative.

Oncocytic Cysts and Cystadenomas

Laryngeal oncocytic cysts and their proliferative counterparts, oncocytic cystadenomas, are uncommonly encountered lesions and have been found in 0.1 to 1% of laryngeal biopsies. The majority occur in the false cords: a compilation of 142 cases revealed the distribution of supraglottic, glottic, and subglottic cases to be 74%, 22%, and 4%, respectively. A female predominance has been noted. Most patients are in their seventh and eighth decades of life. Bilateral and diffuse distribution has been noted, which accounts for the symptomatic recurrence after biopsy. The majority of cases are under 1 cm in greatest dimension. The largest of 19 cases from the AFIP was 2.8 cm in greatest dimension. Rarely, obstructing "extensive" or "bulky" tumors are reported.

Pathology These lesions have acinar, mucinous, or ductal epithelial cells that have undergone oncocytic metaplasia and some degree of hyperplasia. Oncocytic cystadenomas range from predominantly simple cystic lesions (oncocytic cysts) to more complex multicystic and papillocystic lesions (oncocytic cystadenoma). Lymphocytic infiltrates may be prominent, reminiscent of Warthin's tumors. Simple cysts without papillations should be termed oncocytic cysts, and more complex multicystic lesions with a papillary component should be classified as cystadenoma. Atypia is usually absent.

Treatment Simple conservative endoscopic excision is curative for most cases. Occasionally, laryngeal oncocytic cystadenomas may recur, more likely as a manifestation of diffuse or multifocal oncocytic metaplasia, rather than oncologic aggressiveness.

Laryngoceles

The laryngeal ventricle (sinus of Morgagni) is the "pocket" between the vocal fold (true cord) and the ventricular fold (false cord). The bilateral upward extension, or "cul-de-sac," of the ventricle is the laryngeal saccule. A laryngocele is the symptomatic dilation of the laryngeal saccule by entrapped air (i.e., a pulsion diverticulum), which still communicates with the laryngeal lumen. A laryngocele may remain confined to the endolarynx (internal laryngocele) as a supraglottic submucosal bulge. It may also undermine the paraglottic space superiorly, protrude over the superior rim of the thyroid lamina, and herniate through the foramen of the superior laryngeal neurovascular bundle in the thyrohyoid membrane. This type of laryngocele (mixed external/internal, or foramina cyst) presents as an anterior neck mass. It stands to reason that external laryngoceles must have some internal component, and for that reason may be termed "mixed" laryngoceles. Patients with internal and mixed laryngoceles complain of hoarseness, dyspnea, and chronic cough. Newborn infants with laryngoceles present with a feeble cry, difficulty in feeding, cough, and a neck mass. The air-filled nature of laryngoceles an be confirmed radiographically. Sometimes, laryngoceles undergo intermittent obstruction, and secretions will result in a mucus-filled sac. Coughing may clear the obstruction, dispelling the secretions. As long as a communication exists between the sac and the laryngeal lumen, this can still be classified as a laryngocele.

Laryngoceles are usually unilateral, less often bilateral, and may be seen over a wide age range, from neonates to incidental autopsy findings in the middle aged and elderly. Only a small subset of patients with laryngoceles is involved with activities involving increased intralaryngeal pressure (glass blowers, trumpet blowers). It is thought that asymptomatically enlarged saccules may be prevalent in the general population and render persons more vulnerable to laryngocele formation. These enlarged saccules may be a phylogenous laryngeal remnant akin to primate lateral laryngeal air sacs. MacFie radiographically demonstrated a high incidence of asymptomatic laryngoceles (56%) occurring in 93 musicians (wind instrumentalists). These laryngoceles could be demonstrated upon forceful expiration with an open glottis (a maneuver similar to playing a wind instrument) yet could not be demonstrated on forced exhalation with a closed glottis (Valsalva maneuver).

Therapy Laryngoceles are cured by simple excision. Histologically, they are lined by respiratory mucosa. Lymphoid tissue, as the inferior extension of Waldeyer's ring, may also be present. No other neck cyst would present as an air-filled cyst; therefore, this history is pathognomonic.

Saccular Cysts

A saccular cyst is a mucin-filled dilatation of the laryngeal saccule secondary to obstruction, either acquired or congenital in origin analogous to a sinonasal mucocele. "Mucus under pressure points" and saccular cysts may point either medially or laterally. Medial saccular cysts obscure the anterior vocal fold, but they are limited in size and extension by the anterior commissure. Lateral saccular cysts point superolaterally, and like external laryngoceles they may herniate through the thyrohyoid membrane and reach massive proportions if neglected.

Histologically, saccular cysts are lined by saccular mucosa—usually respiratory type, but occasionally squamous or oncocytic mucosa—and filled with mucinous material. This latter point distinguishes saccular cysts from laryngoceles. Saccular cysts may be indistinguishable from thyroglossal duct cysts because remnant thyroid tissue may be absent from the latter. The majority of thyroglossal duct cysts are present in the anterior midline, inferior to the hyoid bone. However, rare thyroglossal duct cysts may push on the thyrohyoid membrane to encroach on the pre-epiglottic space. In this case, the anatomical location aids in distinguishing between the two: the stalk or tract of a thyroglossal duct cyst is midline and leads to the hyoid bone, whereas the stalk of a large saccular cyst is lateral and herniates through the thyrohyoid membrane.

Therapy Saccular cysts and symptomatic mixed laryngoceles may be cured by surgical excision.

Papillomas: Juvenile Onset and Adult Onset Papillomata

Clinical Features Laryngeal papillomas (fungiform papillomas, exophytic papillomas) are benign lesions induced by human papillomavirus (HPV). Juvenile onset laryngeal papillomatosis (JOLP) is distinct from adult onset laryngeal papilloma (AOLP) in that JOLP usually follows an exuberant course. It presents prior to the age of 5, without gender predominance. Papillomas are multiple and may carpet the endolarynx and subglottis resulting in extreme hoarseness and upper airway obstruction. The clinical course of JOLP is often one of innumerable recurrences. Most cases resolve by puberty, but some cases of JOLP may persist into young adulthood. By contrast, AOLP occurs after the second decade of life with a strong male predominance. These lesions are usually singular and amenable to endoscopic excision. Occasionally, AOLP may present with multiple lesions that recur after excision.

Malignant change is known to occur in laryngeal and laryngotracheopulmonary papillomatosis. This change may occur in concert with external promoters, such as irradiation and cigarette smoking, or may happen de novo. Although malignant transformation of JOLP and AOLP has been the subject of sporadic reports, retrospective series place the rate of malignant transformation for all laryngeal papillomata between 2 and 17%. Transformation may occur in localized as well as diffuse cases. Malignant transformation can occur in both JOLP and AOLP and is usually related to long disease duration. A large series of 102 patients with JOLP (52%) and AOLP (48%) revealed that eight patients (7.8%, three with JOLP, five with AOLP) developed malignant transformation. The time between onset of papilloma and diagnosis of carcinoma was 4 to 55 years (mean 24 years). Clinical factors that suggested malignant transformation included decreased vocal fold mobility, the presence of cervical lymph nodes, exuberant and rapid growth requiring very frequent excisions, and laryngeal edema.

Pathology Exophytic papillomas are histologically defined by stratified squamous epithelium over fibrovascular cores. The fibrovascular cores result in their characteristic fungiform architecture and distinguish these lesions from condylomas, which are more sessile and broad based. The squamous mucosa of exophytic papillomas is usually immature without a significant hyperkeratosis, and mild to moderate dysplasia may be present. However, dense keratinization, intramucosal keratinization, diffuse dysplasia of any degree, or full-thickness dysplasia should make one consider squamous carcinoma, ex-fungiform papilloma, or a papillary squamous carcinoma. Increased atypia has been correlated with clinical recurrence, although opinions as to the predictive value of dysplasia in laryngeal papillomas differ.

Prognosis and Treatment As mentioned, patients with JOLP may require innumerable endoscopic laser procedures to maintain airway patency. Interferon therapy may decrease the interval between relapses. AOLP may be treated by conservative endoscopic excision.

Those patients with dysplastic papillomata require close clinical follow-up. In a series of 63 cases of laryngeal papillomas, 12 patients (19%) initially presented as JOLP with disease persistence into adulthood. No case underwent malignant change. Twenty patients (32%) had solitary lesions, cured by endoscopic excision, 30 patients (47%) had multiple lesions, 60% of them required multiple (five or fewer) excisions. Seven patients (8%) developed florid papillomatosis. Aggressive papillomatosis or florid aggressive papillomatosis refers to diffuse laryngotracheal squamous metaplasia and papillomatosis, which carpets the endolarynx and may extend into the tracheobronchial tree and the pulmonary parenchyma. Generally, florid papillomatosis may occur in up to 25% of patients with either JOLP or AOLP. These patients require tracheostomy for airway control and may require laryngectomy for disease control.

Granular Cell Tumors

Granular cell tumors (GCTs) are benign, slow-growing tumors of schwannian origin. There is a female preponderance, and they occur in a greater than expected proportion of African Americans. GCTs occur over a wide age range, with an age peak in the third to fifth decades of life. They are the most common, benign, nonepithelial neoplasm listed in the AFIP Otolaryngic Tumor Registry. About half of all cases involve the head and neck, most commonly the anterior tongue and subcutaneous tissues of the head and neck. The larynx and trachea are less commonly involved, representing 1.6 to 3.7% of involved sites. Other common sites for GCT include the breast, anogenital region, and subcutaneous tissue of the trunk. GCTs rarely grow larger than 3 cm in greatest dimension. Multiple synchronous or metachronous tumors at various sites occur in ~5% of patients.

Laryngeal GCTs are smooth white polypoid tumors arising from the posterior true vocal folds or, less often from the anterior commissure, false cords, subglottis, and trachea. Eighteen cases of pediatric laryngeal GCTs have been reported. Unlike their adult counterparts, they have a predisposition for the anterior subglottis. Tracheal GCTs represented 4% (6 of 145) of referred cases of GCTs to the AFIP; a recent literature review identified 30 tracheal tumors in total. Eighty-four percent of cases occurred in women; 63% of them were African American. Twenty percent of tracheal GCTs were multiple.

Patients with laryngeal GCTs complain usually of hoarseness, those with tracheal tumors invariably have a long history of "intractable asthma." The white mucosal surface is due to squamous mucosal hyperplasia that accompanies about half of these cases.

Pathology GCTs typically have an infiltrative growth pattern and a "histiocytoid" type cytological appearance. The tumor cells grow in small nests and cords; their nuclei are small and generally eccentric, their cytoplasm abundant, granular, or "stippled." They have indistinct cytoplasmic boundaries, and the granules are para-aminosalicylic acid (PAS) positive and resistant to digestion. Nuclear pleomorphism and mitotic figures are not usually seen in GCTs.

Pseudoepitheliomatous hyperplasia may be present in up to 50% of cases; on occasion it may even mimic infiltrating squamous carcinoma (**Fig. 48–1**). Rare cases can have a moderate degree of epithelial atypia. The pseudoepitheliomatous hyperplasia can be a clue, on superficial biopsies, that one may be dealing with a GCT. It should lead the pathologist to look for granular cells in the subepithelial layer or to order deeper sections (**Fig. 48–2**). Marked desmoplasia may be seen in older or larger tumors; here the typical granular cells are seen in the periphery of the tumor. Ultrastructural examination confirms a relationship to Schwann's cells. The cytoplasmic granules are actually lysosomal structures that contain infoldings of cell membranes similar to schwannian extensions. Typically, both GCTs and Schwann's cells express S-100 protein strongly, and may both also express markers of histiocytic differentiation (antibody KP-1, which detects CD 68).

Figure 48–1 *Superficial biopsy of vocal cord granular cell tumor. Note the fingerlike projections of the hyperplastic rete pegs; they can easily mimic infiltrating squamous carcinoma [low power, hematoxylin and eosin (H&E)].*

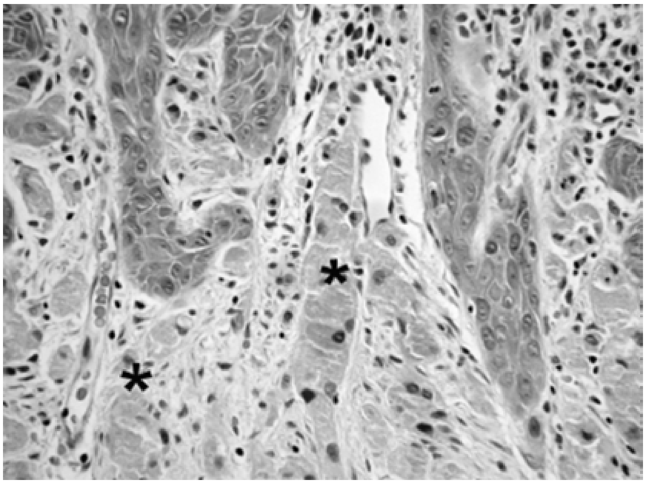


Figure 48–2 *Granular cell tumor. Asterisks marks infiltrating islands of bland histiocytoid granular cells. (High power, H&E).*

Treatment Conservative endoscopic removal will be curative for most cases. GCTs have a very low rate (8%) of recurrence, even after incomplete excision. Recurrent tumors or frankly malignant tumors require resection with free margins.

Laryngeal Paragangliomas

Paraganglia are organs of neural crest origin, situated adjacent to sympathetic or parasympathetic nerves, with capacity for production of an array of neuroendocrine products. In the head and neck, paraganglia are normally present in the middle ear, along the glossopharyngeal and vagus nerves, at the common carotid bifurcation, and along the superior and inferior laryngeal nerves. Head and neck paraganglia are usually parasympathetic, with the exception of those derived from the superior sympathetic cervical ganglia. Head and neck paragangliomas most commonly arise from the carotid bodies (carotid body tumor, chemodectoma), vagus nerve, or middle ear paraganglia ("glomus tumor").

Superior laryngeal paragangliomas outnumber inferior laryngeal paragangliomas with a ratio of ~6:1. Superior laryngeal paragangliomas are polypoid, submucosal intralaryngeal tumors. Inferior laryngeal paragangliomas are usually dumbbell shaped and extend inside and outside the larynx (**Fig. 48–3**). Their hypervascularity imparts a red to blue hue grossly and may result in hemoptysis. Their situation deep to the thyroid fascia may lead to clinical confusion with thyroid tumors. Clinically, paragangliomas of the supraglottis usually present with hoarseness and dyspnea due to a mass effect. The infraglottic tumors may present with dypsnea and hoarseness due to nerve palsy of the recurrent laryngeal nerve. Severe pain has been occasionally noted, presumably due to neuroendocrine activity of the neoplasm. Bleeding may be profuse during biopsy. The female to male ratio of these tumors appears to be 3:1, with a right-sided laryngeal predisposition, with a ratio of 2.3:1. Laryngeal paragangliomas may be associated with paragangliomas elsewhere or with a family history of paragangliomas. Apparent hormonal sensitivity with increased growth during pregnancy has been reported. Rarely, tumors may be clinically functional.

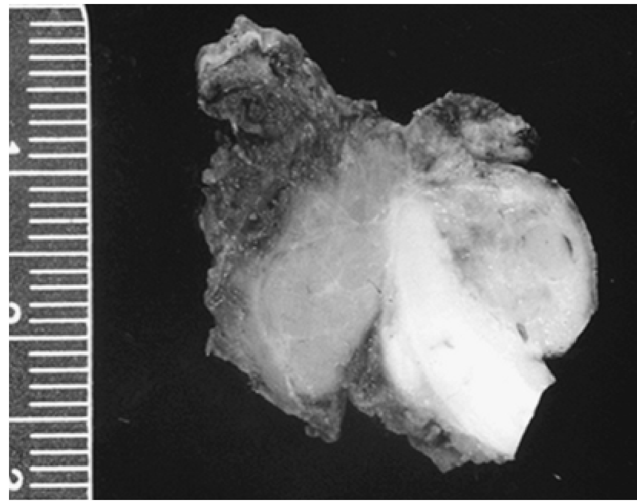


Figure 48–3 *Inferior laryngeal paraganglioma. Note dumbbell shape, as the tumor erodes cricothyroid membrane, producing extralaryngeal and intralaryngeal component.*

Pathology Paragangliomas are vascular, epithelioid neoplasms. Toward the center of the tumor, "balls of cells" (zell-ballen) are formed: the cells are separated into ball-like organoid compartments by fibrovascular tissue (**Fig. 48–4**). This pattern is highlighted by

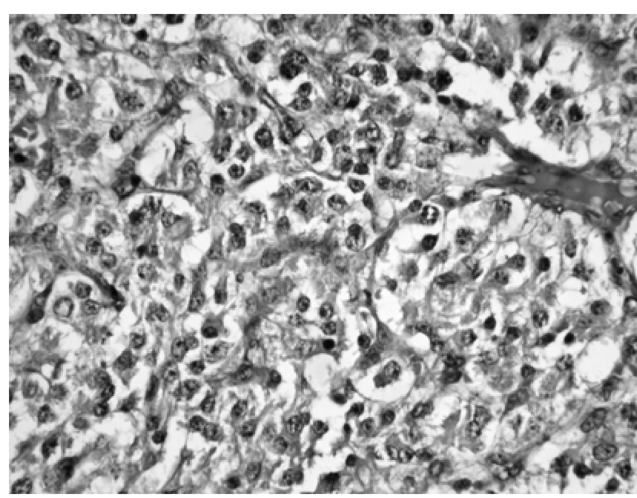


Figure 48–4 *Laryngeal paraganglioma. Note the organoid nest of epithelioid cells with stippled ("salt and pepper") nuclear chromatin pattern (medium power, H&E).*

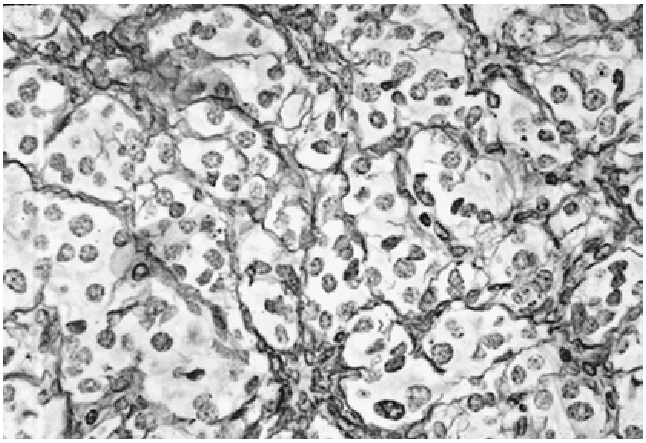


Figure 48–5 *The reticulim stain highlights the nested zell-ballen pattern.*

reticulin stain (**Fig. 48–5**). Cells have abundant granular cytoplasm and round nuclei with "salt and pepper" stippling of chromatin. Cellular pleomorphism can be present in paragangliomas, and occasionally may be marked, but the zell-ballen pattern is maintained at least focally. The pleomorphism is of no independent prognostic value. Paragangliomas are usually positive for neuroendocrine immunohistochemical markers such as neuron-specific enolase, synaptophysin, and chromogranin. The S-100 stain highlights the Schwann-like sustentacular cells that unsheathe the zell-ballen. Paragangliomas are usually negative for epithelial markers such as cytokeratin, carcinoembryonic antigen, and epithelial membrane antigen, and also calcitonin. The lack of calcitonin and epithelial markers may aid in distinguishing paragangliomas from neuroendocrine carcinomas (see later discussion). Ultrastructurally, these tumors contain abundant membrane-bound dense core granules typical of neuroendocrine neoplasia.

Treatment The majority of laryngeal paragangliomas are curable by conservative surgery. Paragangliomas are sensitive to radiotherapy; unresectable skull base paragangliomas may be palliated with radiotherapy. Many so-called reported malignant paragangliomas of the larynx can be better classified as moderately differentiated neuroendocrine carcinomas (NECs). True malignant laryngeal paragangliomas; that is, histologically confirmed paragangliomas that develop metastatic disease, are extremely rare.

Malignant Laryngeal Tumors

Dysplasias

Dysplasia refers to the abnormalities in cellular morphology indicative of premalignant changes. Patients presenting with hoarseness may be found to have varied vocal fold changes, which may be associated with dysplasia. Clinical microlaryngoscopy may reveal white areas (leukoplakia), red patches (erythroplakia), or speckled red and white areas (erythroleukoplakia). Microscopically, leukoplakia corresponds to thickened (hyperplastic) epithelium, coated with a retained keratin layer (hyperkeratosis). This may or may not also contain dysplasia. Microscopically, erythroplakia corresponds to thinned mucosa, usually with submucosal inflammation. The underlying vasculature is closer to the surface, thereby causing a reddened appearance. Erythroplakic lesions are more likely to contain dysplasia, as will be described later. Microscopically, erythroleukoplakia incorporates both of the foregoing histologies.

Dysplasia has classically been graded as mild, moderate, or severe carcinoma in situ (CIS). This schema is based on relative thickness of the atypical cells, defined as those immature cells with nuclear pleomorphism, atypical mitotic figures, and disturbed polarity. These atypical changes are confined to the lower one third of the mucosa for mild dysplasia, extend between the lower one third and upper one third for moderate dysplasia, extend into the upper one third of the mucosa for severe dysplasia, and extend through the full thickness of the mucosa for CIS. Many authors and pathologists will group severe dysplasia and CIS together because they both appear to have equal biological potential to progress to invasive cancer. Other authors have suggested further simplication, grouping dysplasias together as either low grade or high grade. This approach would allow for better classification of histological, "nonclassical" dysplasias, such as those with deep pearls in the rete pegs, or basal cell "teardrop" proliferation, despite retained cellular maturation in the upper mucosa (**Fig. 48–6**).

Clinically, laryngeal CIS has a male predominance and a peak incidence within the sixth decade of life. Patients with CIS may be adequately treated by vocal fold stripping and frequent observation. The incidence of progression to invasive squamous cell carcinoma (cell) varies in most published series of the past two decades, from 25 to 50%. The average latency period of progression is 3 to 5 years.

Squamous Cell Carcinoma

The incidence of laryngeal carcinoma is ~4 per 100,000 population, for all patients (gender, race, age), which has been declining slightly over the last 2 decades. The incidences are higher in males than in females, in nonwhites than in Caucasians, and in those over the sixth

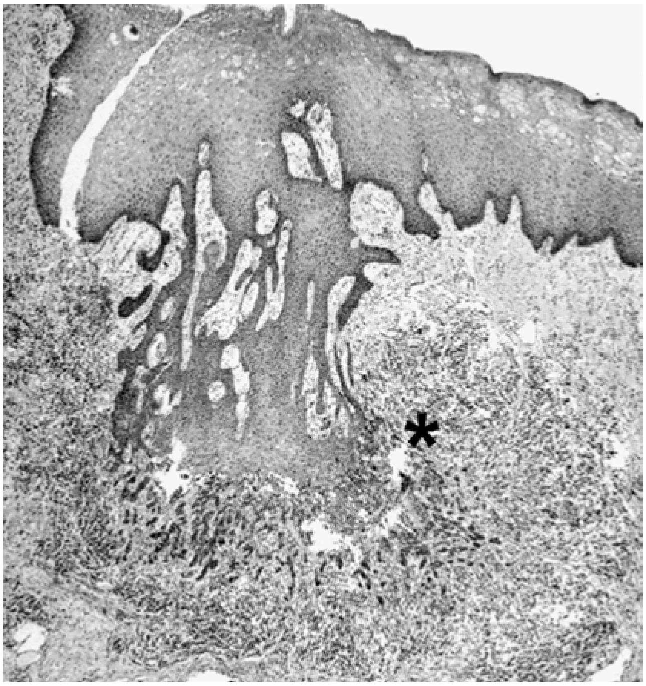


Figure 48–6 *This infiltrating squamous carcinoma defies the standard nosology of dysplasias as progressing upward prior to invasion. Invasive carcinoma (asterisk) develops from hyperplastic rete pegs. However, the overlying mucosa is not dysplastic (low power, H&E).*

decade in life. Squamous cell carcinoma (SCC) comprises more than 95% of laryngeal malignancies. The majority of patients with laryngeal cancer have a history of smoking. Alcohol consumption is more likely associated with supraglottic carcinomas, rather than glottic tumors. Spain, France, and Italy have among the higher world rates of laryngeal cancer; these geographic variations probably reflect dietary factors (alcohol consumption). Other promoting factors include occupational exposures (construction industry, wood working, hydrocarbon exposures) and irradiation.

The larynx can be anatomically divided into three compartments: supraglottic, glottic, and subglottic. The supraglottis is composed of the epiglottis, aryepiglottic folds, vestibular folds (false cords), ventricle, and saccule. The ventricle is the "pocket" between the vocal fold (true cord) and the vestibular fold. The lateral superior extension, or "cul-de-sac," of the ventricle is variably sized and referred to as the saccule. The glottis refers to the vocal folds from the edge of the ventricle to the free edge of the vocal fold. The boundary, which divides the glottic from the infraglottic compartments, is defined as the tissue of the free edge of the vocal fold to the level of the inferior cricoid margin. From a compartmental viewpoint, the supraglottis is distinct and separate from the glottis and subglottis, which are contiguous. The American Joint Committee (AJC) and the International Union Against Cancer (IUAC) classifications will stage tumors on the undersurface of the vocal fold as glottic tumors.

The exact laryngeal site for a tumor may determine or influence (1) the type of presenting symptoms, (2) stage at presentation, (3) surgical options (conservative voice sparing vs radical), and (4) patient prognosis. Although the vast majority of malignancies of the supraglottis and glottis are SCCs, nonsquamous malignancies (e.g., salivary tumors and neuroendocrine carcinoma) are more likely encountered in the supraglottis than in the glottis. Glottic tumors present with changes in voice quality (i.e., hoarseness); patients tend to seek medical care when these tumors are relatively small (1–2 cm). Large glottic tumors or bilateral glottic tumors may present with worsening upper airway obstruction and stridor (**Fig. 48–7**). Supraglottic tumors may reach a larger size before becoming symptomatic. Epiglottic tumors may cause a change in vocal quality (a muffled or "hot potato" voice). Tumors at the base of the epiglottis may be asymptomatic and escape visualization at indirect laryngoscopy ("winklekarzinom" or "cancer in the corner"). The piriform sinuses are extralaryngeal gutters, which flank the thyroid lamina. The lateral wall of the piriform is the thyroid lamina; the medial wall is the cricoid ring. On swallowing, the larynx is raised upward, the epiglottis

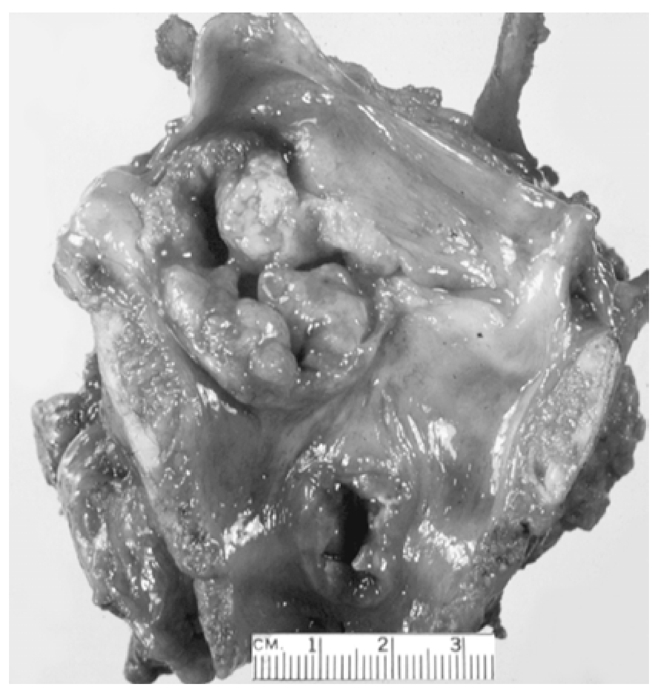


Figure 48–7 *Transglottic squamous carcinoma causing airway obstruction. Note the tracheostomy stoma.*

moves inferiorly, partially covering the endolarynx, and the vocal folds close. Fluids are deflected laterally and inferiorly down the piriform sinuses and are led into the opened cricopharyngeus—the opening to the esophagus. Tumors of the piriform sinus because they are not in the endolarynx, do not present with vocal or respiratory symptoms. They are usually large ulcerating tumors and produce symptoms when they reach considerable size; patients may complain of pain on swallowing that may radiate to the ear.

Most tumors, which appear "infraglottic," actually arise from the undersurface of the vocal fold and so are still considered glottic tumors. Tumors at or below the cricoid may be considered infraglottic tumors (e.g., cricoid chondrosarcomas). These tumors have a more insidious course of onset, with patients complaining of increasing exertional dypsnea. Primary tracheal malignancies are extremely rare; it is more likely to encounter a primary carcinoma of the esophagus eroding into the trachea rather than finding a primary neoplasm.

The supraglottis is derived, embryologically, from the buccopharyngeal anlage (branchial arches 3 and 4), whereas the glottic compartment is derived from the laryngotracheal anlage (arches 5 and 6). The fascial compartmentalization, as well as the lymphatic drainage, is distinct for the supraglottic and glottic compartments. This anatomical fact is the basis for the oncologic soundness of the supraglottic horizontal laryngectomy. Dye injected in the supraglottic space remains confined and does not travel to the ventricular or glottic tissues. Likewise, glottic dye injections do not pass superiorly to the ventricle or inferiorly to the mucosa overlying the cricoid. In fact, the mucosa overlying the lamina propria of this space (Reinke's space or the laryngeal bursa) may burst from fluid distention rather than allowing injected dye to extend into the ventricle or cross the anterior commissure. These studies also confirm that the larynx is divided into right and left compartments.

Two membranes serve as barriers and effect tissue compartmentalization: the quadrangular membrane and the conus elasticus. The quadrangular membrane is present in the supraglottis; the conus elasticus in the glottis and infraglottis both meet and fuse in the area of the perichondrium. These membranes act as a curtain containing tumor medially. The quadrangular membrane originates from the lateral aspects of the epiglottis and extends to the vestibular folds and arytenoid cartilages. Deeper, it fuses with the perichondrium of the thyroid lamina. It continues inferiorly to connect the thyroid cartilage with the cricoid ring (cricothyroid ligament). The conus elasticus (cricovocal membrane) is the continuation of the cricothyroid ligament; it ensheathes the vocalis muscle, separating it from Reinke's space. The conus elasticus merges with the vocalis ligament (vocal tendon). This vocalis ligament serves to limit the spread of carcinoma initially from the vocal fold. A subepithelial periventricular membrane (central membrane) has been identified that spans the paraglottic region and, in effect, connects the conus elasticus and quadrangular membranes centrally.

The vocalis muscles originate from the vocal ligaments, which have their attachment at the anterior commissure and insert upon the arytenoid processes. The attachment of the vocal ligaments also limits the spread of carcinoma from one lateral side to the other. The vocal folds have a relative paucity of lymphatics, as compared with the supraglottis and the pre-epiglottic space. This paucity of lymphatic vessels is most marked in the anterior vocal folds and accounts for the rarity of cervical metastases of T1 glottic carcinomas. The lymphatic channels of the vocal fold become denser posteriorly in the region of the arytenoids. Glottic carcinomas may spread by undermining the tissue around the ventricles (paraglottic region), escaping the endolarynx and spreading laterally by invading the cricothyroid ligament and the inferior aspect of the thyroid lamina. Ossified thyroid lamina is more prone to tumor invasion as compared with the relatively avascular nonossified cartilage. Carcinoma may also spread superiorly into the vestibular fold by undermining paraglottic ventricular tissue.

The epiglottis is composed of fenestrated cartilage, which allows for early tumor spread from the laryngeal surface to the lingual surface and into the pre-epiglottic space. The latter contains abundant lymphatics; tumor spread into this space increases the risk for cervical metastasis and worsens prognosis. The pre-epiglottic space is bound anteriorly by the thyrohyoid membrane. Tumor that breaches this space invades into the base of tongue. The superior boundary of the pre-epiglottic space is the hyoepiglottic ligament, which connects the hyoid bone to the epiglottis. Epiglottic carcinomas, which are inferior to the hyoepiglottic ligament (infrahyoid tumors), are more commonly encountered than those superior to the hyoepiglottic ligament (suprahyoid tumors). The hyoepiglottic ligament provides a barrier blocking the inferior passage of the infrequent suprahyoid carcinomas into the pre-epiglottic space.

Histologically, SCCs may be recognized by their ability to produce keratin (**Fig. 48–8**). Intercellular bridges may be seen as fine hairlike structures between cells. Well- and moderately differentiated SCCs are usually associated with a surface mucosal component (in situ carcinoma), which clinically appears as an erythroleukoplakic irregularities. Poorly differentiated SCCs may

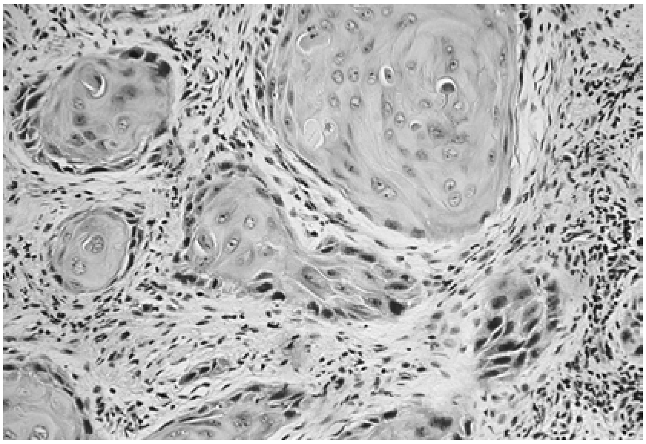


Figure 48–8 *Irregular infiltrating islands of squamous carcinoma producing keratin pearls (high power, H&E).*

appear clinically ulcerated or entirely submucosal, with little perceptible mucosal involvement. Two variants of SCC deserve mention: verrucous carcinoma and spindle carcinoma.

Verrucous carcinoma (VC) is a clinicopathologically distinct, well-differentiated variant of squamous carcinoma, which is cytologically benign yet clinically aggressive. VC presents as a slow-growing, gray-white, firm, warty tumor with a "cauliflower" like surface and sharply demarcated margins (**Fig. 48–9**). Clinically, lymph nodes are usually palpable in patients with VCs that are benign but reactive. A male predominance is seen with VC, which relates to cigarette-smoking and tobacco-chewing habits. The oral cavity is the most common site for VC, usually on the buccal mucosa or gingival; the larynx is a less common site of occurrence.

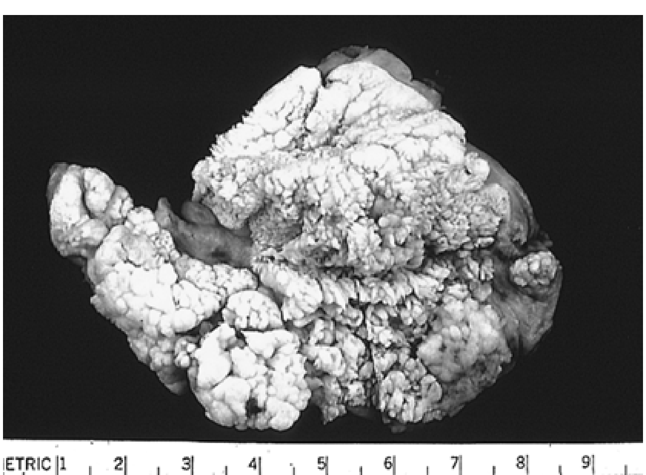


Figure 48–9 *Buccal verrucous carcinoma, with typical cauliflower-like gray-white papillary surface.*

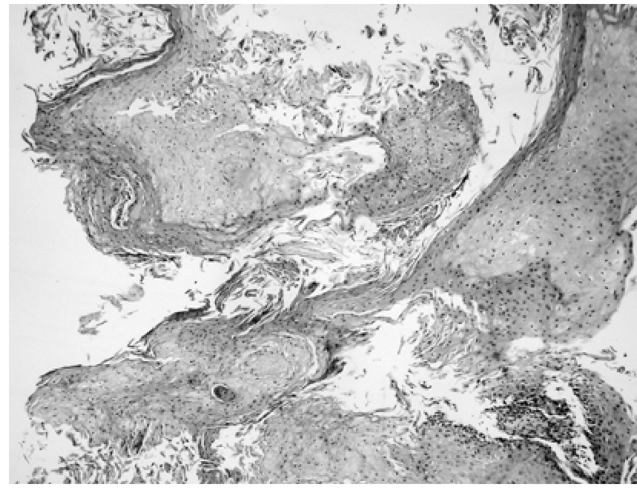


Figure 48–10 *Superficial biopsy of verrucous carcinoma reveals only nondysplastic keratinous debris (low power, H&E).*

The clinical and histological similarities of VC to the "Buschke-Lowenstein" giant condyloma of the genital tract served to fuel interest in the association of VC and HPV. HPV sequences (mostly HPV 18) have been detected in 20 to 100% of patients with oral or laryngeal VC by polymerase chain reaction plus Southern blot hybridization. Superficial biopsies of VC may be quite misleading because they may yield no more than keratinous debris (**Fig. 48–10**). VC is characterized by a papillary hyperkeratotic and parakeratotic surface component. The deep portion of VC reveals pushing, broad, anastomosing rete pegs (**Fig. 48–11**). A band of chronic inflammation accompanies the pushing margins in about half the cases. Adjacent to the tumor, one sees mucosal hyperplasia and rete pegs that are elongated and anastomosing, yet still thin. Nuclear pleomorphism is rarely seen in this tumor. When typical squamous carcinoma develops in a VC, it can be differentiated from VC by

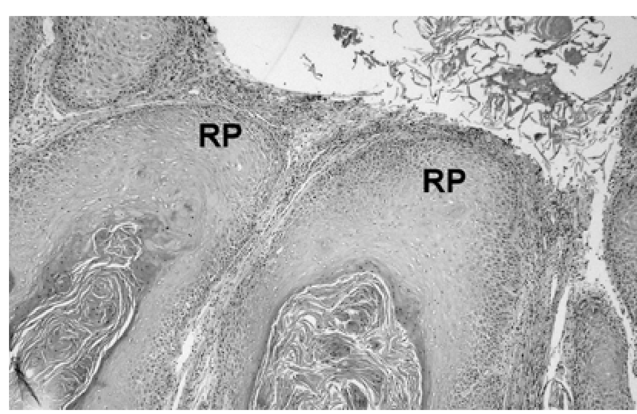


Figure 48–11 *The diagnosis of verrucous carcinoma is confirmed by the finding of pushing bulbous rete pegs (RP) (medium power, H&E).*

nuclear pleomorphism and irregularly shaped islands of infiltration, distinct from the broad pushing margins of VC.

VC is primarily a surgical disease. Resection with adequate margins is usually curative. However, sometimes anatomical location results in constraints on tumor resectability. Irradiation is acceptable therapy in these cases, albeit with a significant recurrence rate. Squamous carcinoma may arise in a VC, referred to as a "hybrid tumor." Irradiation may promote the emergence of higher grade neoplasia; however, squamous carcinoma can occur in VC without previous irradiation. Furthermore, not all irradiated patients with VC develop "anaplastic degeneration." A large series on 104 oral tumors reported by Medina and associates has shown that careful sampling of VC may reveal hybrid tumors in up to 20% of cases, unrelated to previous irradiation. When patients with VC are resected, the surgeons should refrain from radical neck dissection. If evidence of squamous carcinoma is identified after adequate sampling, then the issue of elective neck dissection is reevaluated in this light.

Spindle squamous cell carcinomas (SpCCs) may be associated with a prominent malignant spindle cell component with a wide spectrum of appearances; the general term *spindle cell carcinomas* has been advocated for these tumors, which may cause diagnostic problems on preoperative biopsies. SpCCs are uncommon; they represented 0.6% (12 of 2052) of laryngeal malignancies seen by Ferlito (1976). Yet for all upper aerodigestive tract SpCCs, the larynx and hypopharynx are common sites; 65% of these cases occurred in the larynx/epiglottis/vocal cords/pyriformsand hypopharynx. There is a pronounced male predisposition, and most patients are between the fifth and ninth decades. The tumors may be polypoid and exophytic, or ulcerating and infiltrating, but the tendency for laryngeal tumors is to retain an exophytic polypoid growth pattern. Histologically, SpCCs contain either in situ SCC or invasive SCC, with an additional, obviously malignant spindle cell component. This latter malignant component is to be distinguished from the reactive fibroblastic population commonly seen adjacent to invasive neoplasia. The malignant spindle cell component may express mesenchymal features by light microscopy (e.g., chondroid or osteoid production), immunohistochemistry (e.g., muscle markers), or electron microscopy (intermediate or contractile filaments). However, these spindle cells may also express keratin, belying their squamous origins.

Wide resection is indicated; the role of adjuvant chemotherapy and radiotherapy is uncertain. Generally, polypoid exophytic tumors have an improved prognosis over invasive and ulcerating tumors as a function of presenting stage. Hellquist and Olofsson reported only one tumor-related death (at 2 years), and 12 patients were disease free, five of them for 5 years or longer. Olsen and colleagues (1997) reported several interesting findings from their series of 34 patients with laryngeal (25) and hypopharyngeal (nine) SpCC. Patients with laryngeal SpCC had an improved 3-year survival (76.2%), as compared with those with hypopharyngeal SpCC (56.8%), which probably relates to inherent resectability of tumors (especially polypoid ones) confined to the larynx. Keratin expression was significantly related to decreased survival, which probably is a function of the inherent capacity for lymphatic metastasis of the carcinomatous component.

Neuroendocrine Tumors

Neuroendocrine tumors of the larynx are divided into two broad categories based on their tissue of origin: epithelial and paraganglionic. The epithelial-derived tumors, known as neuroendocrine carcinomas (NECs), are uncommon neoplasms constituting 0.06% of laryngeal malignancies. Due to differences in biological behavior and histologic growth patterns, this group of neoplasms is further subclassified into three distinct subtypes: carcinoid tumor [well-differentiated neuroendocrine carcinoma (WDNEC)], atypical carcinoid tumor [moderately differentiated neuroendocrine carcinoma (MDNEC)], and small cell carcinoma, including both the intermediate and oat cell variants [poorly differentiated neuroendocrine carcinoma (PDNEC)]. Patients with WDNEC survive longer with less morbidity than with PDNEC, and MDNEC has a biological behavior intermediate between these two.

MDNECs are the most frequently encountered type of NEC, followed by PDNEC. True, "typical," carcinoid WDNECs are the least common subtype and are extremely rare. A recent review from the AFIP indicated a ratio of 54:14:2 for MDNEC, PDNEC, and WDNEC, respectively, out of 8469 malignant laryngeal neoplasms. Patients most often present in the sixth to eighth decade of life. There is a strong male predisposition. The most frequent presenting symptom is hoarseness. Patients with PDNECs often present with a neck mass. There is a strong association of MDNEC and PDNEC with a history of smoking. Tumors most commonly arise in supraglottic sites with only very occasional tumors arising in other areas of the larynx (**Fig. 48–12**).

Patients present with submucosal or polypoid masses usually ranging in size from a few millimeters up to 4 cm in greatest dimension. WDNECs are characterized by nests of uniform cells separated by a fibrovascular or hyalinized connective tissue stroma. Nuclei are round to

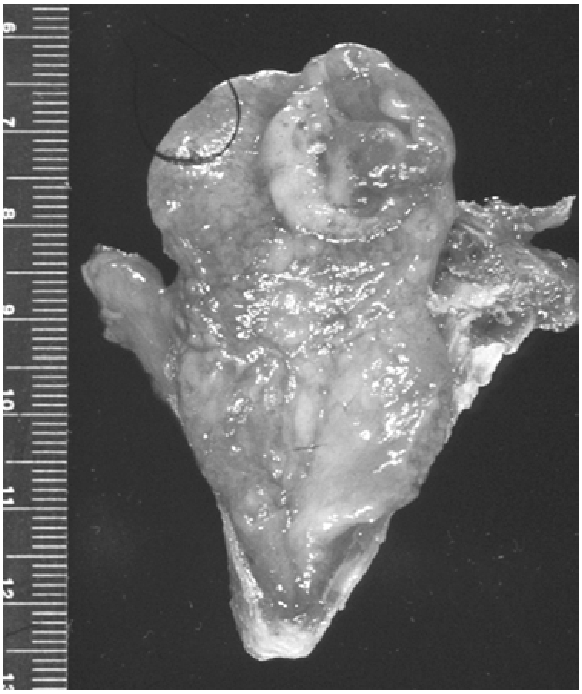


Figure 48–12 *Supraglottic ulcerated laryngeal neuroendocrine carcinoma.*

oval with stippled or vesicular chromatin and eosinophilic cytoplasm. A glandular component is common. Cellular pleomorphism, mitotic activity, or necrosis is usually absent in WDNEC. MDNECs are characterized by infiltrative growth and a varied histologic pattern, which may include glandular, organoid, acinar, trabecular, solid, and nesting architectures (**Fig. 48–13**). The tumor cells are polyhedral to round, are at least twice the size of the small cell variant of PDNEC, and contain varying amounts of eosinophilic cytoplasm. Occasionally, oncocytic differentiation may be observed. Nuclei are round to oval with a stippled chromatin, are often eccentrically located, and display mild to severe pleomorphism. Mitotic figures are rare or absent. PDNECs are characterized by sheets and, rarely, interconnecting ribbons of undifferentiated small cells with minimal cytoplasm (oat cell variant), or slightly larger cells with minimal to moderate cytoplasm (intermediate variant) and hyperchromatic, pleomorphic, oval, round, or spindle-shaped nuclei with delicate chromatin and absent or inconspicuous nucleoli. Individual cell necrosis, vascular and perineural invasion, and prominent mitotic activity are common. Glandular and/or squamous differentiation and rosette formation may occasionally be observed.

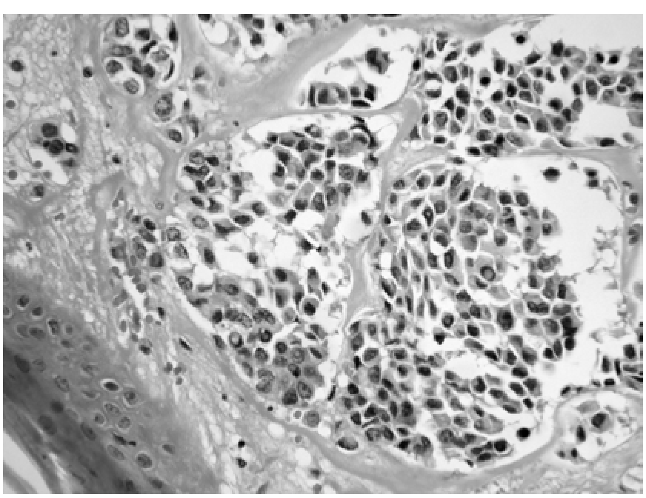


Figure 48–13 *This moderately differentiated laryngeal neuroendocrine carcinoma has a paraganglioma-like organoid pattern and stippled nuclear chromatin. Unlike paraganglioma, the profile of immunohistochemical expression will include epithelial markers and calcitonin (high power, H&E).*

Ultrastructural studies demonstrate neurosecretory granules in all three types in varying numbers and sizes; desmosomes and tonofilaments are frequent in WDNEC and MDNEC and are less common in PDNEC, whereas lumina (true and intracellular) are frequent in WDNEC and MDNEC and are usually absent in PDNEC. Argyrophil silver stains are frequently positive in WDNEC and MDNEC and are usually negative or focally positive in PDNEC. Classic neuroendocrine markers (chromogranin, neuron-specific enolase, synaptophysin, Leu-7, etc.) will usually be positive. Calcitonin is also positive in all three types of neuroendocrine carcinoma.

Surgery is the primary therapy for WDNEC and MDNEC, and a lymph node dissection is indicated for MDNECs due to their high rate of cervical nodal metastases. PDNECs should be treated with a combination of radiation and chemotherapy similar to the protocols used for pulmonary oat cell carcinoma of the lung due to early hematogenous spread. WDNECs have a very good prognosis. In a recent review, eight of 12 patients were disease-free 1.5 to 8 years after treatment; only one patient had died of disease after 5 years. MDNECs are more aggressive neoplasms, with 5- and 10-year cumulative survival rates of 48 and 30%, respectively. Tumors larger than 1 cm appear to be more aggressive, and patients developing skin or subcutaneous involvement have a worse prognosis. In a recent review of 119 MDNECs with follow-up information, 74% of the 66 patients treated with neck dissections during their disease course had metastatic disease. PDNECs are the most aggressive type of laryngeal NEC. In a recent series 73% of patients with PDNEC died with an average survival of only 9.8 months (range 1–26 months). Two- and 5-year survivals were 16 and 5%, respectively.

Adenoid Cystic Carcinoma

Salivary gland–type neoplasms are rare tumors in the larynx, accounting for less than 0.7% of laryngeal carcinomas. They account for 56% of laryngeal glandular tumors; malignancies outnumber benign tumors by a ratio of 2.6:1. The most common benign tumors are oncocytic lesions (see oncocytic cystadenomas already discussed); benign mixed tumors are a distant second, unlike their frequent occurrence in the salivary glands. The four most common malignant salivary tumors are adenosquamous carcinoma, followed with equal incidences of adenoid cystic and mucoepidermoid carcinoma, and lastly malignant mixed tumors. Other salivary tumors that arise, albeit rarely, in the larynx include myoepithelioma (benign and malignant), acinic cell carcinoma, epithelial myoepithelial carcinoma, clear cell carcinoma, and salivary duct carcinoma.

Adenoid cystic carcinoma (ACC) of the larynx represents from 0.07 to 0.25% of laryngeal carcinomas. To date, ~125 cases have been reported. There is a broad age range of occurrence, with a slightly increased incidence in the fourth to sixth decades of life. Approximately 60% involve the subglottis, 33% the supraglottis, and 6% the vocal fold. Voice change or hoarseness, pain radiating to the ear, and dysphagia are the most common presenting symptoms for supraglottic tumors; subglottic tumors may be associated with "asthma," pain, hoarseness, or dyspnea on exertion. Extralaryngeal invasion may result in initial presentation as a thyroid mass.

The majority of tumors diffusely invade the submucosa and adjacent soft tissues without protruding into the laryngeal lumen (**Fig. 48–14**). ACC may form three patterns: tubular, cribriform, and solid. The cribriform is the most frequent and the solid pattern the least frequent pattern observed. Adenoid cystic carcinomas are composed of cells of two types: ductal cells and abluminal myoepithelial cells. The ductal cells are bland population with oval basophilic nuclei with homogeneous chromatin distribution, and usually with little cytoplasm, reminiscent of basal cell carcinoma of the skin. The nuclei are frequently angulated and may rarely have coarse chromatin and prominent nucleoli. Although these latter two features are more likely in the solid, high-grade tumors, high-grade cytology may occasionally be seen with intermediate-grade tumors.

A mixture of patterns is common; classification and hence grading are made according to the predominant pattern. If a tumor has more than 30% of the solid pattern, it is classified as the solid variant, which is grade III. The tubular pattern (well differentiated or grade I) is characterized by slender tubules, solid cords, and glandular structures infiltrating a well-hyalinized background. These are composed of myoepithelial cells, often surrounding central luminal–forming epithelial structures. The cribriform pattern (moderately differentiated or grade II) is characterized by invasive tumor islands with multiple "holes" (pseudocysts or pseudolumina) punched out in a "Swiss cheese" or sievelike pattern (**Fig. 48–15**). The pseudolumina are sharply demarcated from the surrounding cells and may contain a "rind" of dense pink basement membrane material and central blue mucopolysaccharides, or they may be entirely filled by the basement membrane material. The solid pattern (poorly differentiated or grade III) consists

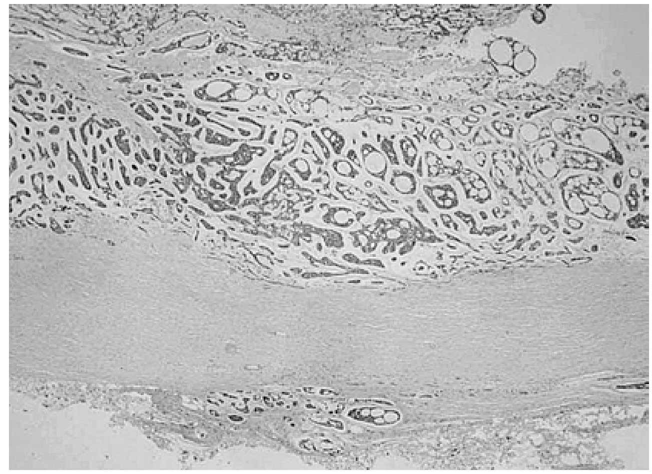


Figure 48–14 *Infiltrating adenoid cystic carcinoma (low power, H&E).*

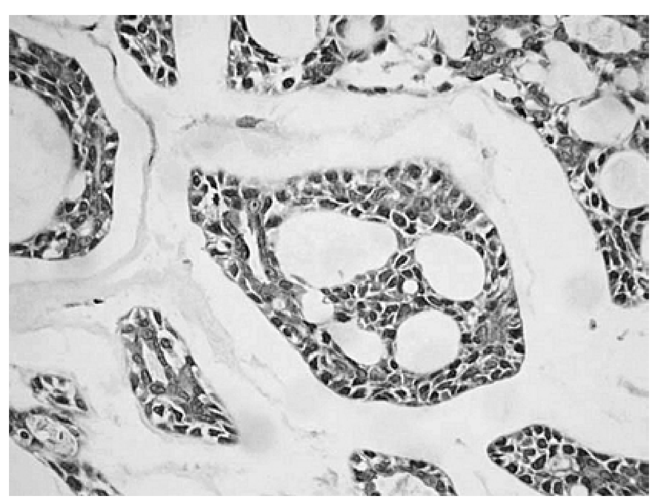


Figure 48–15 *Adenoid cystic carcinoma. Note islands with "sievelike" sharply demarcated pseudoglands (high power, H&E).*

of large islands of carcinoma composed predominantly of myoepithelial cells with infrequent true lumina lined by cuboidal epithelial cells, with only occasional punctuation by pseudocysts. Perineural spread is a feature common to all patterns. Mitotic figures and apoptotic cells are occasionally present in intermediate-grade tumors and are common to high-grade tumors. Necrosis is seen, usually only in the solid pattern.

Complete surgical excision consisting of partial or complete laryngectomy is the treatment of choice. Recurrence rates historically have been in the range of 50%; therefore, adjuvant radiation therapy is indicated. Cervical lymph node dissection usually is not recommended unless there is palpable lymphadenopathy. Isolated pulmonary metastasis, if present, should be treated aggressively and not alter treatment of the primary tumor. ACC has a slow, relentless clinical course marked by high local recurrence rates and late distant metastases. Patients with subglottic tumors have an average survival of 8 years, with occasional patients surviving for 15 years. Tumors with perineural invasion appear to have a higher risk of local recurrence. Occasional patients may have rapid progression of disease or transition from a slow to a more rapid clinical course. Similar to ACC arising in other head and neck sites, most patients will, unfortunately, eventually die from their tumor.

Chondrosarcoma

Laryngeal chondrosarcomas occur mostly in hyaline cartilage structures: the ratio of cricoid to thyroid lamina tumors is ~3:1. Cricoid tumors often occur in the posterior cricoid ring and present with slowly evolving dyspnea. Thyroid lamina tumors grow anterolaterally and tend to present earlier as palpable neck masses, which might clinically mimic thyroid tumors. Chondrosarcomas occur infrequently in the epiglottis, which is composed of elastic cartilage and remains nonossified, or the arytenoid cartilages, which undergo ossification inconsistently. The hyoid bone ossifies early in life, and chondrosarcomas in this site are extremely uncommon, although hyoid bone chondrosarcoma has been reported as part of the spectrum of Gardner's syndrome.

The peak incidence of laryngeal chondrosarcomas occurs when cartilage ossification is more likely to be present,:during the fifth to eighth decades. The site preference for chondrosarcomas (posterior and posterolateral cricoid, and inferior-lateral thyroid lamina) corresponds to areas of laryngeal muscle insertion. The diagnosis may be established by preoperative radiograph that can reveal an expansile tumor of the cricoid or thyroid alae with calcifications; however, some tumors may not reveal calcifications.

The gross appearance of laryngeal chondrosarcomas is characteristic, appearing as expansile tumors with a glassy firm white-gray cut surface. Tumoral calcification lends a gritty feel to the tumor. Laryngeal chondrosarcomas follow the diagnostic criteria of chondrosarcomas of the axial skeleton. Grade I tumors grow in a lobulated fashion with abundant cartilage tissue. Crowding of chondrocytes and double nuclei within single lacunae are seen. Cytologically, grade I tumors have minimal atypia, but nucleoli, not normally seen in chondrocytes, are present. Mitotic figures are usually not seen. Grade II tumors have greater cellularity and more nuclear crowding than grade I tumors. Obvious cytological pleomorphism and hyperchromatism and mitotic activity can be appreciated. Grade III tumors begin to lose their chondroblastic differentiation: elaboration of chondroid material is sparser and solid areas of malignant cells are present; there is brisk mitotic activity.

Uncommonly, chondrosarcomas can progress to develop a histologically high-grade "dedifferentiated" component, referred to as additional malignant mesenchymal component (AMMC); this can occur in up to 10% of pelvic chondrosarcomas. A few cases of dedifferentiated laryngeal chondrosarcomas have been reported. Grossly, the dedifferentiated component loses its glassy cartilaginous appearance and has a softer tan-gray component that may "spring forth" from the tumor. The significance of an additional malignant mesenchymal component is that the dedifferentiated component generally augers poor patient survival. Tumors with an additional mesenchymal component histologically appear as a spindle cell sarcoma that was sharply demarcated from the grade I or chondrosarcoma. The spindle component differs from a grade III chondrosarcoma in that there is no tendency toward histologic cartilage differentiation.

Laryngeal chondrosarcomas generally have an indolent nature because the majority of these tumors are either grade I or II. In the larynx, experience dictates that it is wiser to treat all cartilaginous tumors as potential chondrosarcomas in that they require complete resection with negative margins. Recurrence is very rare if the tumor is completely resected. "Shelling out" even a small "chondroma" will guarantee eventual recurrence. This usually occurs after a few years, but it may take up to a decade or longer. Recurrence may "convert" an initial potentially curative plan of partial laryngectomy to salvage a total laryngectomy after recurrence. The literature supports the idea that recurrent laryngeal chondrosarcomas generally do not result in a compromised patient outcome: a metastatic rate of 8.5% has been estimated from reported cases, which correlates with histologic grade.

Palma and colleagues culled a handful of hyoid bone chondrosarcomas from the European literature and noted a high tendency for recurrence. It may be that the marrow space of the hyoid is quite prone to harbor satellite foci of tumor. It would appear reasonable to recommend removing the entire hyoid bone, rather than partial hyoid resection.

Dedifferentiated laryngeal chondrosarcomas appear to "live up" to their ominous reputation, although they have rarely been reported, and cases with clinical follow-up are even rarer. We reported a patient with laryngeal dedifferentiated chondrosarcoma who died of disease 17 years after the first onset of chondrosarcoma, 3 years after the onset of the dedifferentiated component. Ferlito et al's patient died 30 months after total laryngectomy. Nakayama and colleagues described two patients: one was convincingly reported and illustrated as having a progression from a grade I to a higher grade of dedifferentiated chondrosarcoma; this patient was alive with persistent local and metastatic disease.

Osteogenic Sarcoma

Laryngeal osteosarcoma (LOS) is a more rarely encountered entity than laryngeal chondrosarcoma (LCS). Although over 300 cases of LCS have been reported in the literature, and many more cases have gone unreported, few cases of LOS have been identified. Unlike LCS, which expands out from the cricoid ring or thyroid lamina, LOS does not occur primarily in the bony/cartilaginous framework but as a polypoid soft tissue tumor of the endolarynx. Whereas LCS is usually associated with an indolent history of progressive dyspnea, patients with LOS manifest rapid onset of symptoms matching the tumor's rapid growth. Pinsolle and coauthors tabulated seven cases of LOS, six of which were pathologically documented. These six cases occurred in males in the seventh and eighth decades of life.

Grossly, these tumors grow as fleshy submucosal polyps, mimicking other sarcomas. Calcification may be noted grossly. The laryngeal cartilaginous/bony framework is usually noted to be intact. Microscopically, LOSs are high-grade neoplasms with either a fibrosarcomatous or an osteoblastic osteosarcoma appearance. Malignant stellate or spindled sarcoma cells are seen with a variable component of osteoid. The latter can form a delicate eosinophilic latticework pattern, or a denser, well-formed osteoid matrix. Chondroid areas can be seen but are not predominant. Osteoclastic multinucleated giant cells are frequently found in LOS; they are usually not observed in LCS.

Complete surgical resection with adequate margins is indicated for LOS. Given its general high morbidity, conservative surgery with plans of reconstruction is probably not warranted. Five of the six reported patients died from metastatic disease 3 months to 2 years after diagnosis (mean 1 year); one patient was alive after 40 months. The distinction between LCO and LOS has important prognostic implications because LCS is generally associated with a low mortality due to low tumor grade. Curiously, the two aforementioned cases of LCS associated with osteoid formation were associated with aggressive metastatic behavior similar to LCO. However, the behavior of these two cases is probably related to overall tumor stage.

Liposarcomas

Liposarcoma is one of the most common soft tissue sarcomas of adulthood, usually occurring in the lower extremities and in the retroperitoneum. The head and neck are involved in 5.6% of liposarcomas, usually the soft tissues of the neck, scalp, and face. Hypopharyngeal and laryngeal sites occurred in 38% of the 76 cases of head and neck liposarcomas reviewed from the Royal Marsden Hospital over a 50-year period. A pronounced male predominance is noted: only two reported cases occurred in females. There is a wide age range reported from the third decade of life onward, with a supraglottic predisposition. Tumors may extensively involve the hypopharynx, piriform sinuses, and supraglottic and/or glottic compartments, causing progressively increasing airway obstruction and/or vocal changes. Liposarcomas appear as submucosal polypoid pedunculated tumors that are soft and yellow-tan-gray. They may be as small as 2 cm or be massive and transglottic. Submucosal tumors may present endoscopically as small bulges, with overlying edematous yet benign mucosa, rendering superficial biopsies nondiagnostic. As with liposarcomas at other sites, occasional laryngeal tumors may be part of a multicentric clinical picture: one case has been reported of an obese man with a previous myxoid liposarcoma of the thigh who developed a higher-grade supraglottic liposarcoma.

Hypopharyngeal/laryngeal liposarcomas of the larynx reflect the spectrum of histology seen in the skeletal soft tissues. Tumors may be low grade (lipoblastic liposarcoma or lipoma-like, sclerosing liposarcoma or atypical lipoma) or intermediate to high grade (myxoid liposarcoma, pleomorphic liposarcoma, round cell liposarcoma, dedifferentiated liposarcoma). Wenig and colleagues (1988) noted that the sclerosing pattern was the most common one encountered for hypopharyngeal/laryngeal tumors. Low-grade tumors are characterized by an abundance of mature, histologically benign adipose tissue, coursed by collagenous fibrous tissue. Lipoblasts may be focal; they have characteristic "chicken claw"–shaped nuclei that are indented by cytoplasmic fat globules. Their chromatin is usually dense and pyknotic, but enlarged nucleoli may be found. Atypical lipoblasts, which have large, irregular nuclei and smudged chromatin, and florette cells, with multiple nuclei in a wreathlike pattern, can be seen. An abundant collagenous, bland fibroblastic background may dominate the picture; hence the tumor may be classified as sclerosing liposarcoma. The same histological picture in the subcutis may be diagnosed as an atypical lipoma.

Myxoid liposarcoma is a common histologic pattern generally seen in soft tissue liposarcoma. The stroma is loose and myxoid, perforated by a fine "chicken wire" meshwork of arborizing vessels. The lipoblasts appear as univacuolated signet ring cells and multivacuolated cells. The lipoblasts may be scarce, congregated at the periphery of the expanding tumor lobules. The poorly differentiated forms of liposarcoma may appear as pleomorphic, round cell, or dedifferentiated forms. Pleomorphic liposarcomas are characterized by densely packed malignant spindle cells and bizarre, highly pleomorphic forms; the lipoblastic component (lipoblasts and signet ring cells) may be minimal. Round cell liposarcoma is the "small cell" version of this sarcoma, composed of closely packed "signet ring" type lipoblasts with little intervening myxoid or adipose stroma.

Benign lipomas of the hypopharynx and larynx are rare and must be distinguished from well-differentiated liposarcomas; this may be difficult on limited biopsy of a fatty tumor. It is always possible that features of well-differentiated liposarcoma lurk beyond the preoperative slide. Unfortunately, because the liposarcomas may be small (2 cm in diameter) and lipomas may be enormous, clinical correlation may be of no use. Benign fatty tumors are rare in the endolarynx and hypopharynx; hypopharyngeal lipomas are more common than lipomas of the intrinsic larynx, where there is a predisposition for the supraglottic structures. As with liposarcomas, these tumors may be clinically silent and reach large proportions before coming to diagnosis. In the hypopharynx, lipomas may form long, pedunculated, sausagelike tumors, causing progressive dysphagia with solid foods. These polypoid masses can prolapse into the endolarynx, causing airway obstruction, or into the oral cavity, resulting in gagging.

Lipomas are well-circumscribed, noninfiltrative tumors composed of mature adipose tissue and fine fibrous septate. True lipoblasts and atypia are not present. The adipose tissue may be admixed with other mature, benign mesenchymal tissue, necessitating diagnostic subclassification; that is, tumors with somewhat thicker collagenous septae and bland fibroblasts in addition to the mature adipose tissue are fibrolipomata. Laryngeal spindle cell lipomata have denser infiltrates of bland fibroblastic-like spindle cells in addition to collagen bundles. Intramuscular lipoma show interspersed bundles of mature skeletal muscle, enveloped and surrounded by mature adipose tissue. Myxolipomata have a prominent myxoid background.

Liposarcomas are properly treated by resection with adequate margins. Wenig et al (1988) argued that, whereas tumors of similar low-grade histology may be called "atypical lipomatous tumors" of the subcutaneous or intramuscular tissues, this may encourage inadequate removal in the upper aerodigestive tract. Conservative, function-sparing yet curative partial laryngectomy is preferable to salvage total laryngectomy in the face of an inadequately treated recurrent tumor. The majority of reported tumors developed single or multiple recurrences, after initial polypectomy or subtotal resection and/or high tumor grade. Occasional tumors may develop higher-grade clones as they recur. Radical neck dissection usually is not warranted for low-grade tumors. High-grade tumors may develop locoregional metastases. Adjuvant radiotherapy may be indicated for high-grade tumors. The 5-year survival for laryngeal liposarcomas (89%) is significantly better than for some other head and neck sites, such as soft tissue of the neck (60%), pharynx (59%), and oral cavity (50%). This relates to the inherent resectability of laryngeal tumors, as well as a predisposition for laryngeal liposarcomas to be low grade (ratio of low to high grade = 2.3) as compared with other head and neck sites (ratio of low to high grade = 1.2 for neck and pharynx).

THE ORAL CAVITY AND PHARYNX

Benign Tumors

Mucoceles

Mucoceles (also referred to as retention cysts and extravasation cysts) represent the most commonly encountered salivary pathology. Mucoceles are the result of obstruction, cystic dilation, and rupture of minor salivary gland ducts with mucus extravasation into adjacent soft tissue. Antecedent trauma or obstruction is requisite: either calculus, trauma, inflammation, postsurgical complication (from stenosis or ligature of a duct), or synchronous tumor mass. Mucoceles usually present in the first three decades of life, but they may be encountered at any age. Any site with minor salivary tissue can become involved, the most common being the lower lip. Clinically, they appear as painless, raised, translucent to bluish white, soft masses that can fluctuate in size. Superficial lesions appear vesicular; deeper lesions are covered by normal-appearing mucosa. Major salivary mucoceles occur most often in the submandibular gland. Sublingual mucoceles are termed ranulas, which can be subclassified as either simple or plunging. The elevated mucosa of a ranula often has a characteristic bluish color, and the mucosal distension is reminiscent of the appearance of the gullet pouch of a frog. A simple ranula, which is more common, remains confined to the floor of the mouth above the level of the mylohyoid muscle. A plunging or deep ranula is a ruptured sublingual mucocele that usually extends below the level of the mylohyoid muscle and occasionally into the mediastinum. Ranulas are notable for producing larger, more noticeable swellings than minor salivary mucoceles, but typically they cause minimal discomfort. They rarely can lead to airway compromise. Clinically, cervical ranulas may simulate cystic hygroma and other potential midline submental-neck lesions such as dermoid cyst, lymphadenopathy, and hematoma. Plunging ranulas appear to occur with greater incidence in the Maori and Pacific Island/Polynesian populations. The precise etiology of their predisposition is unknown, although local trauma or inherent mylohyoid dehiscences may play important roles.

Histologically, mucoceles can be lined by cuboidal, columnar, or squamous epithelium. If the dilated ductal cyst has ruptured, then the actual cyst wall (pseudocyst) is composed of granulation tissue. Foamy histiocytes are a prominent reaction to secretions. Variable amounts of eosinophilic extracellular mucin are present and can be highlighted on mucicarmine stain. The adjacent salivary gland can appear atrophic and inflamed. Treatment is excision of the mucocele and the surrounding minor salivary glands. The mucocele may recur if the suspect gland is not completely removed.

Fibromas, Submucous Fibrosis, and Fibromatoses

The term *fibroma*, within the oral cavity, refers to a reactive fibroblastic process, usually secondary to mechanical irritation such as ill-fitting dentures or biting. Lesions are usually single, occurring most commonly at buccal, lingual, or labial sites. They appear as submucosal, sessile, or pedunculated masses, which can have hyperplastic, papillary surfaces. Fibromas at specific sites have acquired their own terminology: those of the palate, which occur secondary to dentures, are called "leaf-shaped" fibroma, those of the gingiva are called epulis fissurata, and those occurring adjacent to the mandibular cuspid are called retrocuspid papilla. Diffuse or papillary fibrous overgrowth may occur in the gingiva (diffuse gingival hyperplasia), which may be hereditary and associated with numerous constitutional syndromes (e.g., Laband's syndrome, Cowden's syndrome, tuberous sclerosis, Ramon's syndrome). Acquired gingival hyperplasia can be seen as a complication to numerous drugs, such as phenytoin, bleomycin, cyclosporine, and verapamil.

Histologically, a fibroma is composed of bland fibroblastic population with collagen deposition. The overlying mucosa may be atrophic or quite hyperplastic to the point of being papillary. A significant giant cell component also may be present (giant cell fibroma). Simple excision is curative for fibromas. Diffuse gingival hyperplasia may be treated by gingivectomy. Acquired gingival hyperplasia may regress with discontinuation of the causative drug.

Oral submucous fibrosis is a reactive response to chewing betel quid, a practice especially prevalent in India. Rather than causing a mass, submucous fibrosis presents clinically as a diffusely mottled, atrophic, stiffened mucosa, which progresses to cause trismus. Microscopically, submucous fibrosis appears as hyalinized submucosal fibrosis and chronic inflammatory infiltrate. The mucosa is atrophic and can be dysplastic.

Oral submucous fibrosis is irreversible, even after cessation of betel quid chewing. Suspicious areas require biopsies because 10% of patients with this condition progress to developing squamous carcinoma.

The term *fibromatosis* (desmoid tumor, desmoid fibromatosis, or grade I fibrosarcoma) refers to a neoplastic, nonreactive, process. Generally, fibromatoses form irregular, infiltrative, progressively enlarging masses. They usually manifest as mandibular lesions and will show evidence of bony erosion or invasion. Histologically, fibromatosis appears as fibroblastic infiltrate, with infiltrating margins and variable cellularity. If cellular pleomorphic, mitotic figures or necroses are seen, then the diagnosis of grade II or grade III fibrosarcoma may be warranted. Resection with negative margins is the treatment of choice.

Lingual Thyroid

Embryologically, the thyroid anlage develops from both central and lateral components. The central component arises as a midline diverticulum of the developing pharynx, late in the fourth week of development. This component descends and gives rise to the thyroid isthmus and central parts of each lateral lobe. The path of descent, commencing at the fifth week of development, forms the thyroglossal duct, which can be traced from the tongue base (foramen cecum) to the isthmus. Sometimes it can be seen as a persistent pyramidal lobe. Although the track usually becomes obliterated, its persistence is not uncommon and can give rise to cystic dilation (thyroglossal duct cyst). Residual thyroid tissue may present within a thyroglossal duct cyst and can give rise to additional thyroid pathology. The lateral aspects of the thyroid lobes are derived from the lateral thyroid anlage, which arises from the fourth and fifth branchial cleft pouches. The ultimobranchial bodies develop from these pouches and contain migrating neural crest cells. These bodies migrate medially and form the lateral aspects of the thyroid lobes; they also contribute to the neuroendocrine component of the thyroid (the calcitonin-producing parafollicular C cells).

This complicated pattern of thyroid migration allows for numerous developmental variations. Failures of descent of the medial thyroid anlage will result in a lingual thyroid. This is the most common site for thyroid ectopia and may represent a patient's only functioning thyroid tissue. There is a pronounced female predominance. Lingual thyroid may be asymptomatic or may cause dysphagia. Symptoms may worsen with hormonal fluctuations such as menses and pregnancy. Lingual thyroid appears clinically as a smooth, submucosal mass. Sudden growth may cause an ulcerated appearance, mimicking malignancy.

Histologically, lingual thyroid appears as normal thyroid; however, insinuation of this choristomatous tissue between lingual muscle bundles may mimic malignant infiltration. Likewise, we have seen angiomatous change within an ulcerated lingual thyroid mimicking angiosarcoma. Pathology within a lingual thyroid can mirror that of the thyroid gland; thus goiterous change, adenomas, or papillary carcinomas can be seen.

Suggested Readings

Ackerman L. Verrucous carcinoma of the oral cavity. Surgery 1948;23:670–678

Barnes L. Paraganglioma of the larynx: a critical review of the literature. Otorhinolaryngol and Relat Spec 1991;53:220–234

Batsakis JG, El-Naggar AK, Luna MA. Thyroid gland ectopias. Ann Otol Rhinol Laryngol 1996;105(12):996–1000

Brandwein MS, Huvos AG. Laryngeal oncocytic cystadenomas: a report of eight cases and a literature review. Arch Otolaryngol Head Neck Surg 1995;121:1302–1305

Brandwein M, Moore S, Som P, Biller H. Laryngeal chondrosarcomas: a clinicopathologic study of 11 cases including two chondrosarcomas with additional malignant mesenchymal component. Laryngoscope 1992;8:858–867

DeSanto LW, Devine KD, Weiland LH. Cysts of the larynx: classification. Laryngoscope 1970;80:145–176

Ferlito A. Histologic classification of larynx and hypopharynx cancers and their clinical implications: pathologic aspects of 2052 malignant neoplasms diagnosed at the ORL Department of Padua University from 1966 to 1976. Acta Otolaryngol Stockh 1976;342(Suppl):1–88

Lack EE, Worja GF, Callihan LD, et al. Granular cell tumor: a clinicopathologic study of 110 patients. J Surg Oncol 1980;13:301–306

Neel HB, Unni KK. Cartilaginous tumors of the larynx: a series of 33 patients. Otolaryngol Head Neck Surg 1982;90:201–207

Newman BH, Taxy JB, Laker HI. Laryngeal cysts in adults: a clinicopathologic study of 20 cases. Am J Clin Pathol 1984; 81:715–720

Olsen KD, Lewis JE, Suman VJ. Spindle cell carcinoma of the larynx and hypopharynx. Otolaryngol Head Neck Surg 1997;116:47–52

Wenig BM, Hyams VJ, Heffner DK. Moderately differentiated neuroendocrine carcinoma of the larynx: a clinicopathologic study of 54 cases. Cancer 1988;62:2658–2676

Wenig BM, Weiss SW, Gnepp DR. Laryngeal and hypopharyngeal liposarcoma: a clinicopathologic study of 10 cases with a comparison to soft-tissue counterparts. Am J Surg Pathol 1990;14:134–141

Woodruff JM, Senie RT. Atypical carcinoid tumor of the larynx: a critical review of the literature. Otol Rhinol Laryngol 1991;53:194–209

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. Which of the following entities may lead to a false diagnosis of laryngeal squamous carcinoma?
   - A. Granular cell tumor
   - B. Blastomycosis
   - C. Radiation changes
   - D. All of the above
   - E. Granular cell tumor and radiation changes
2. Which of the following statements is true?
   - A. Laryngeal paragangliomas are inherently more aggressive than paragangliomata arising from elsewhere.
   - B. Moderately differentiated laryngeal neuroendocrine carcinoma is uniformly fatal.

C. Most laryngeal paragangliomas arise from the inferior paraganglia.
D. Laryngeal neuroendocrine carcinoma may express calcitonin.

3. Laryngeal chondrosarcomas
   A. Are usually low grade
   B. Most commonly arise from the anterior commissure
   C. Commonly metastasize to the lungs
   D. Can be adequately treated by curettage

4. Ectopic thyroid
   A. Should always be resected
   B. Symptomatology may follow hormonal fluctuations
   C. Most commonly occurs in the lateral cervical compartment
   D. Should always be considered as having malignant potential

Chapter 49

ORIGINS AND SPECIFICATION OF CRANIOFACIAL MUSCULOSKELETAL TISSUES

DREW M. NODEN

The mammalian head develops as a result of epithelial tubes becoming surrounded by mesenchymal cells. The epithelial structures are the brain, located dorsally, and the pharynx (later including the stomodeum), located ventrally. At later stages, additional epithelial structures such as the nasal pits, optic vesicles, laryngotracheal diverticulum, and otic vesicles also develop. The two mesenchymal populations are mesoderm and the neural crest. Both arise in restricted locations and their members then often undergo long-range movements, during which period mesodermal and neural crest cells interact with nearby epithelial tissues as well as with each other.

These movements and interactions culminate in the formation of the musculoskeletal system of the head. Abnormalities in the movements, growth, or differentiation of these mesenchymal populations, due either to genetic or teratogenic insults that target them or adjacent epithelial tissues, account for the majority of craniofacial developmental defects. This chapter presents the origins, movements, and fates of craniofacial mesenchymal populations and discusses some of the interactions necessary for their proper development.

HEAD MESODERM POPULATIONS

Three mesodermal populations plus the notochord form during gastrulation. Prechordal mesoderm is a sparse mesenchymal population established immediately rostral to the notochord, beneath the prosencephalic neural plate. In avian embryos, these cells contribute to the extraocular muscles innervated by the oculomotor nerve; no mapping data are available in mammalian embryos. Prechordal mesoderm cells are an essential source of signals necessary for the bilateralization of the rostral neural plate. Some of the genetic mutations that result in severe holoprosencephaly are deficits in these signals; also, it is believed that teratogenic insult to this population underlies milder forms of midline defects such as fetal alcohol syndrome.

Lateral mesoderm loosely surrounds the pharynx. Ventral to the pharynx, many of these cells participate in the formation of the heart, especially the outflow tract, and also form angioblasts that contribute to many head vessels. At the level of the caudal pharynx, lateral mesoderm forms the cricoid and arytenoid cartilages and other connective tissues associated with the larynx and trachea.

Head paraxial mesoderm is generated during gastrulation as the primitive streak elongates. It is, with respect to position and fates, very similar to somites, which are segmented epithelial condensations of paraxial mesoderm found from the level of the otic vesicle to the tail of the embryo. However, head paraxial mesoderm does not form epithelial structures and lacks many of the transcription factors normally found in emerging and newly formed somites (e.g., *scleraxis*, *paraxis*, *pax1*). Some investigators have reported the presence in head paraxial mesoderm of slight transverse indentations and focal superficial swirls that define indistinct iterative domains called somitomeres. Although this may reflect a vestigial segmental organization important in early vertebrates, there is no cellular or molecular evidence of segmentation in head paraxial mesoderm.

PARAXIAL MESODERM

Paraxial mesoderm forms most of the connective tissues that surround the midbrain and hindbrain (**Figs. 49–1** and **49–2**). These include the early chondrocranium caudal to the pituitary and later the endochondral and intramembranous bones that encase and protect the brain and inner ear. The caudal part of the skull is derived from occipital somites, which develop much like vertebrae except they lack persistent segmentation.

All the skeletal muscles of the head (and trunk) arise from cells located in paraxial mesoderm. The prechordal myoblasts mentioned above move into paraxial locations during the early neurula stage and become indistinguishable from paraxial mesoderm. As shown in **Fig. 49–1**, precursors of head muscles arise in a rostrocaudal order that parallels the site at which motor nerves innervating them exit the brain. There is also a mediolateral separation, with the extraocular muscles generally arising deep in the head, close to the brain, and the branchial muscles arising more superficially.

Beginning caudal to the ear, precursors of the laryngeal and tongue muscles arise in the ventrolateral part of occipital somitic myotomes. These cells break away from the somitic epithelium and migrate ventrally as the hypoglossal cord, moving around and then ventral to the pharynx. Some of the signals promoting these movements are identical to those used by somitic myoblasts that move into the developing limbs.

Head muscles use the same set of myogenic-commitment transcription factors as do trunk muscles, including activation of *myf5* and *myoD*, although these

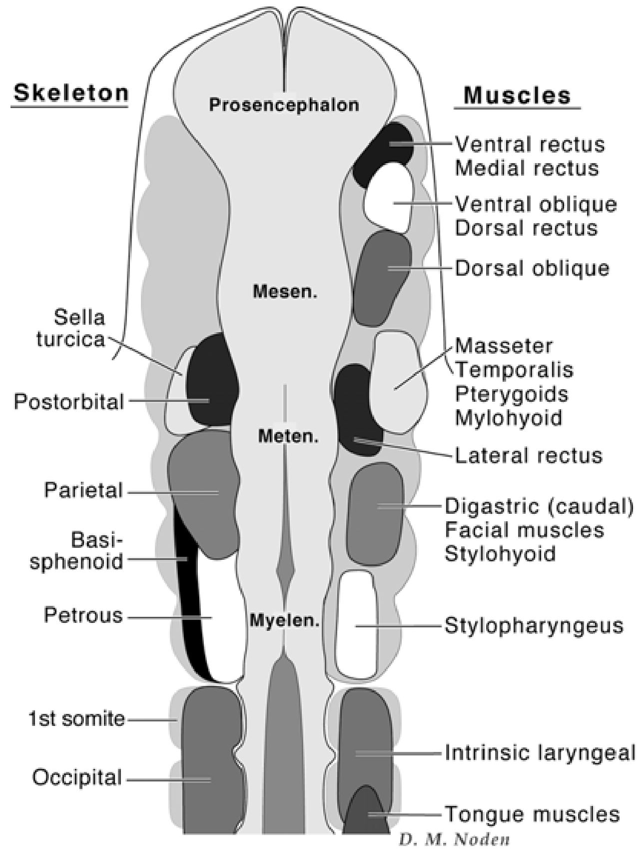


Figure 49–1 *Schematic dorsal view of a neurula-stage (seven-somite) embryo showing the locations of muscle precursors (right side) and skeletal precursors (left) in head paraxial mesoderm. The sella turcica develops at the crest:mesoderm interface and receives contributions from both; all connective tissues rostral to this location are derived from the neural crest.*

are activated using distinct and separate promoter sites. Most muscle precursors initiate expression of these regulatory genes as they move from their sites of origin, shown in **Fig. 49–1**, to their definitive locations in branchial arch and periocular regions, shown in **Fig. 49–3**. During these movements, muscle progenitors move out of the paraxial mesoderm environment and enter neural crest–derived mesenchymal tissue. Their differentiation as multinucleated myotubes and formation of tendinous attachments occurs entirely within the neural crest environment. The only exceptions are the intrinsic laryngeal muscles, which break away from the hypoglossal cord and move into a lateral mesoderm environment beside the caudal pharynx.

Paraxial and lateral mesoderms are the source of all craniofacial endothelial cells. Shortly after the formation of these mesoderms, cells commit to the angiogenic lineage and undergo extensive, invasive movements,

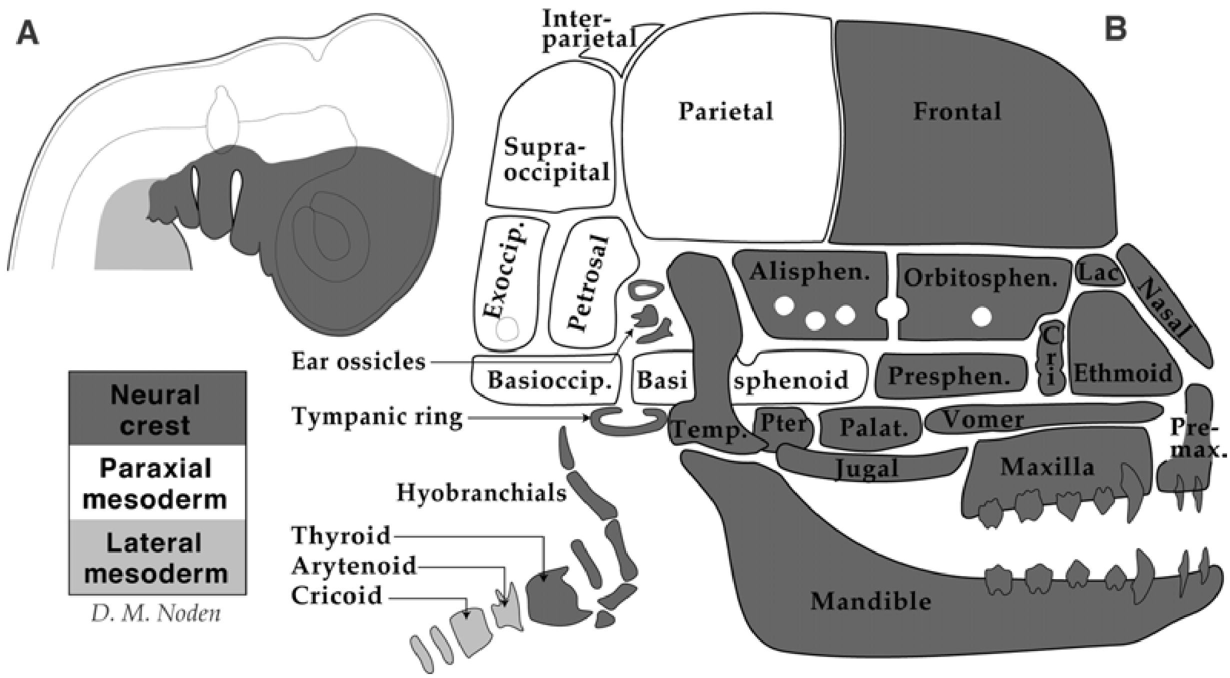


Figure 49–2 ***(A)*** *The interface between connective tissue precursors derived from neural crest, lateral mesoderm, and paraxial mesoderm. The locations of these interfaces have remained constant throughout vertebrate evolution.* ***(B)*** *The skeletal tissues derived from each of these three precursor populations in mammals. Identifying precisely the avian and mammalian homologues involving parts of the postorbital chondrocranium, the sphenoid complex, and the otic capsule is not yet possible.*

which bring them into all parts of the head. These angioblasts are the only truly invasive population of cells in the embryo, and their ubiquitous, systemic movements ensure that all parts of the embryo are seeded with endothelial precursors.

MOVEMENTS AND FATES OF NEURAL CREST

Neural crest cells arise from the neural folds, beginning in most mammals prior to the elevation and fusion of the cephalic folds. Head crest cells are generated in a burst rather than over a protracted period, as is the case in the trunk. The rostral limit of crest generation is the caudal prosencephlalon. After closure of the rostral neuropore and during the early stages of optic vesicle outgrowth, the crest population overlying the mesencephalon spreads rostrally around the prosencephalon as well as laterally, over the optic vesicle and paraxial mesoderm immediately caudal to it, as shown schematically in **Fig. 49–4**. Only the most lateral (future anterior) part of the optic vesicle is not contacted by crest cells because their movement into the future lens and corneal regions is blocked by fibrous adhesions between the vesicle and overlying surface ectoderm. Later, after the lens is formed, crest cells invade this region to form the posterior epithelium and stroma of the cornea.

Crest cells surround and remain in close juxtaposition to the prosencephalon and form all the connective tissues that encompass the forebrain and eyes. Thus, as the telencephalic hemispheres later expand and overgrow the midbrain and part of the cerebellum, the crest-derived connective tissue precursors accompany them. Recently, the use of transgenic mice in which all neural crest cells carry a genetic label has shown that the entire frontal bone is of neural crest origin in this species. This contrasts with the situation found in avian embryos, in which the frontal bone is chimeric, with the supraorbital region derived from the neural crest but the large component forming the roof of the braincase derived from mesoderm. This does not represent a fundamental difference in embryonic origins, but rather reflects the fact that multiple ossification centers arise and fuse to form each of the intramembranous bones of the calvaria. The avian frontal bone is developmentally identical to the mammalian frontal and parietal, and the avian parietal more closely corresponds to the mammalian interparietal.

Another clinically important element of mammalian neural crest development is its contribution to the formation of sutures, which are unique zones of bone

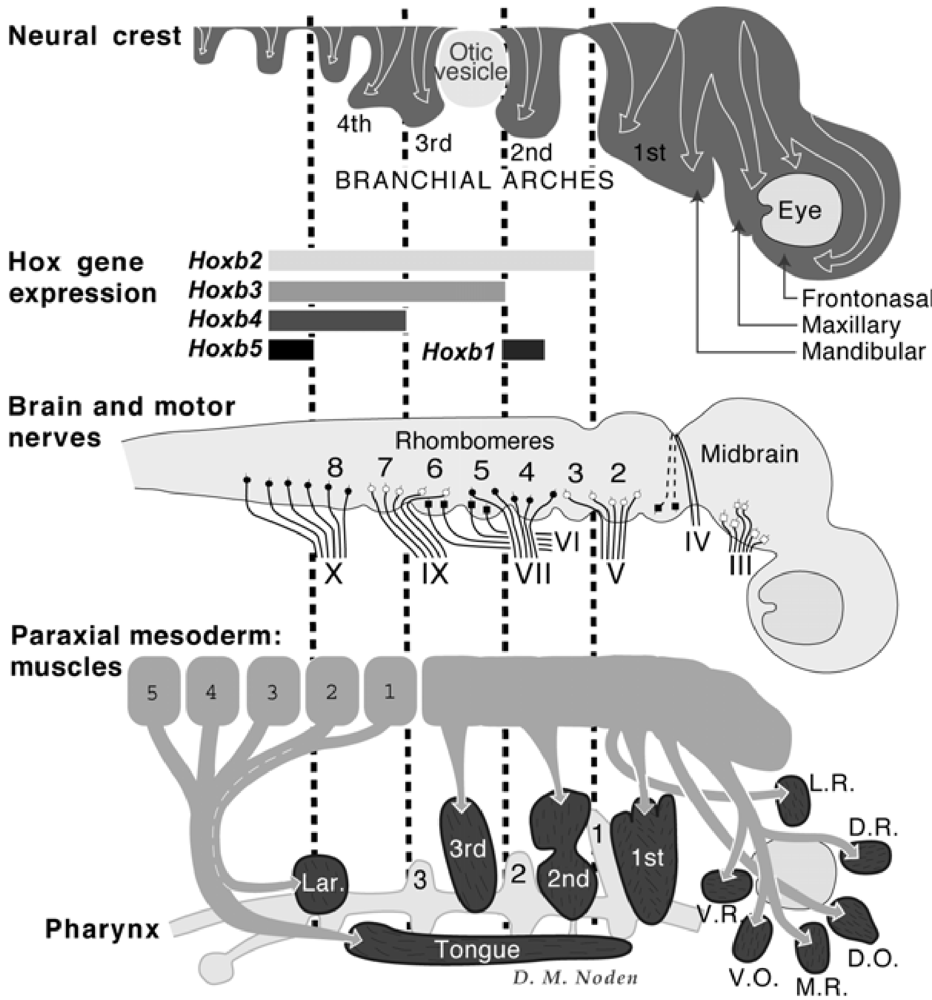


Figure 49–3 The movements of neural crest and myogenic mesoderm cells and the sites of motor axon emergence from the brain. Progenitors for each branchial arch arise at the same axial level and remain in registration throughout their development. Neural crest cells associated with each of these branchial arches express distinct combinations of Hox transcription factors. In contrast, precursors of periocular muscles, connective tissues, and motor axons arise from separate axial levels and do not come into close association until after their migrations/outgrowths are completed.

overlap and sustained proliferation that are essential for the continued growth of the skull roof needed to accommodate brain enlargement. Crest cells are the source of mesenchymal cells in sutures not only associated with frontal bones (metotic and coronal sutures), but also the parietals (sagittal suture), even though the rest of the ossified regions of the parietal bones are of mesodermal origin.

Neural crest cells from the midbrain and hindbrain are the source of all connective tissues found in branchial (pharyngeal) arches. Crest populations destined to form these arches are punctuated by zones of crest cell degeneration that partially segregate progenitors of each arch, and this separation is reinforced by migration-restricting signals emanating from paraxial mesoderm and invaginating otic epithelium.

The migratory routes taken by crest cells to branchial arches 1, 2, and 3 parallel those taken by migrating myoblasts associated with these arches. Moreover, these routes are identical to those taken by motor axons as they emerge from the brainstem. These are displayed together in **Fig. 49–3**. Thus, whereas each of these cell types originates in a different tissue, all arise at the same axial level and remain in registration as they move peripherally to form branchial arches. This continuous nearest-neighbor relationship allows for extensive interactions among all neuromusculoskeletal progenitors for each branchial arch. In contrast, the extraocular muscles arise from quite different axial levels than their connective tissues or innervating axons, and close relations among these three are not established until after the migrations of each are completed.

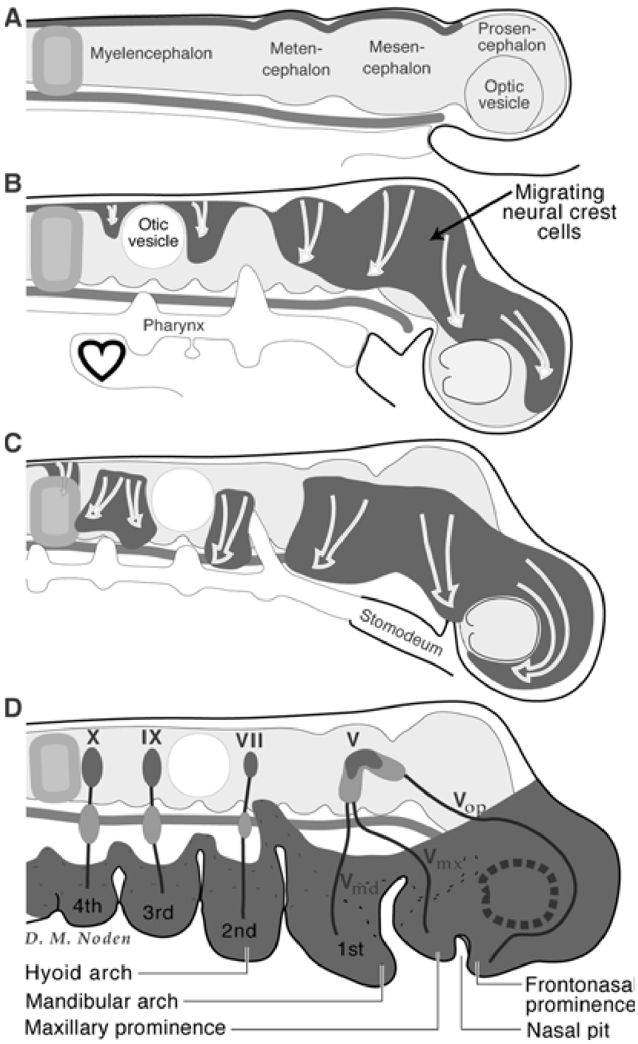


Figure 49–4 *Summary of the pathways taken by neural crest cells arising at different axial locations. Note especially that most neural crest cells fully leave their locations of origin and move into branchial arches, stopping only to contribute to cranial sensory and autonomic ganglia. Neural crest cells associated with the prosencephalon are an exception to this. Here, crest cells fully envelop the brain and remain closely apposed as meningeal tissues, the frontonasal skeleton, and parts of the maxillary prominences.*

Registration among branchial progenitors is lost caudal to the third branchial arch. Although likely present in earlier vertebrates, this region has undergone such extensive evolutionary change that the primitive segmental organization is lost. Crest cells arising from the occipital level of the hindbrain are the source of several unique and essential derivatives: neurons and glia of the enteric nervous system and connective tissues of the distal cardiac outflow tract, where their presence is essential for the normal separation of aortic from pulmonary arteries.

TABLE 49–1 SHARED AND UNIQUE DERIVATIVES OF NEURAL CREST AND MESODERM

Mesoderm Only	Mesoderm and Neural Crest	Neural Crest Only
Skeletal muscle	Dermal bone	Skull sutures
Cardiac muscle	Endochondral bone	Sensory neurons
Endothelium	Cartilage	Autonomic neurons
Blood	Dermis	Enteric neurons
Intestinal smooth muscle	Perivascular smooth muscle	Schwann's cells
Gonad	Tendons	Corneal stroma
Kidney	Ligaments	Odontoblasts
	Pericytes	Melanocytes
	Meninges	C cells (thyroid)

PARAXIAL MESODERM AND NEURAL CREST: WHY BOTH ARE NEEDED

As summarized in **Table 49–1**, both neural crest and mesodermal populations form most types of connective tissues, differing in the locations but not the cellular or molecular features of these tissues. Both also give rise to some derivatives not shared by the other.

Throughout the body, the locations at which initial skeletal condensations form and the shapes they quickly establish are based on an interplay between intrinsic properties of skeletogenic mesenchyme and signals emanating from surrounding tissues. Neural crest cells grafted into somites or early limb buds will form cartilaginous nodules, but they are unable to form normally shaped vertebrae or limb bones; nor can mesoderm form normal facial skeletal structures.

Moreover, crest cells that emigrate from different regions of the brain often have restricted morphogenetic capabilities. For example, replacing second branchial arch crest precursors with first arch precursors results in the formation of a jaw skeleton, appropriate for the first arch, in this ectopic second arch location. The patterns by which branchial arch crest cells respond to skeletogenic stimuli is based on information they acquire prior to their migrations. The genetic basis for this lies in part in the rostral expression boundaries of transcription factors produced by *Hox* genes, some of which are illustrated in **Fig. 49–3**. Elimination of *HoxB2* removes the second arch identity from crest cells that migrate into this arch, and they instead form a first arch skeleton.

In these embryos, the musculatory and vascular systems are also transformed into a first arch anatomical array, even though their precursors were not directly affected by the transplantation or gene knockout. Thus, in the branchial arches as well as other parts of the body, connective tissue precursors are atop the pattern-generating hierarchy, with myogenic and angiogenic populations subordinate.

This neural crest–driven patterning does not hold true for the midface and periocular regions. Here, signals from the prosencephalon and surface ectoderm dictate to crest cells the sites and patterns of skeletogenesis. Clinically, this is manifest in the many facial dysmorphologies associated with varying degrees of holoprosencephaly.

SUMMARY

No single model for partitioning the developing head can encompass both the many regionally autonomous and region-interdependent processes that underly normal craniofacial morphogenesis. Anatomists often consider the branchial and midfacial regions as distinct and separate compartments, but during development there are obligatory influxes of cells from the neural crest and paraxial mesoderm into these regions. These movements bring not only unique cell lineages such as myoblasts and angioblasts into all parts of the head, but also, in the branchial arches, connective tissue precursors endowed with essential pattern-generating capability.

Suggested Readings

Jatiang X, Iseki S, Maxson RE, Sucov HM, Morriss-Kay GM. Tissue origins and interactions in the mammalian skull vault. Dev Biol 2002;241:106–116

Noden DM. Origins and assembly of avian embryonic blood vessels. Ann NY Acad Sci 1990;588:236–249

Noden DM. The role of the neural crest in patterning avian cranial skeletal, connective and muscle tissues. Dev Biol 1983;96: 144–165

Noden DM. Vertebrate craniofacial development: the relation between ontogenetic process and morphological outcome. Brain Behav Evol 1991;38:190–225

Roessler E, Muenke M. Midline and laterality defects: left and right meet in the middle. Bioessays 2001;23:888–900

Trainor PA, Krumlauf R. *Hox* genes, neural crest cells and branchial arch patterning. Curr Opin Cell Biol 2001;13:698–705

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. Connective tissues of the head originate developmentally from the
   A. Neural crest
   B. Paraxial mesoderm
   C. Lateral mesoderm
   D. All the above
2. Development of the midface differs from that in branchial arches in which way?
   A. There are no neural crest cells in the midface region.
   B. Neural crest cells are the key pattern-generating population in the midface.
   C. Neural crest cells are subordinate to the prosencephalon in pattern generation.
   D. None of the above
3. Periocular skeletal muscles are formed by
   A. Migrating neural crest cells
   B. Mesoderm cells that move in registration with migrating neural crest cells
   C. Mesoderm cells that move independently of migrating neural crest cells
   D. Stationary mesoderm cells that originate in the sites where they differentiate

Chapter 50

SURGICAL ANATOMY OF THE NECK AND CLASSIFICATION OF DISSECTIONS

Richard V. Smith and Dorothy Frenz

Frequently, the anatomy of the neck is one of the major factors that attract physicians to the specialty of otolaryngology. The intricate relationships of muscles, nerves, veins, and arteries make this area extremely fascinating. A detailed knowledge of the anatomical relationships of these structures to each other, and to the surface anatomy is required to adequately understand the medical and surgical implications of diseases affecting the neck. The surgical anatomy of the neck is often based on different principles than those taught in most medical school anatomy classes. Although the general concept of triangles of the neck holds true, the clinical distinction of these triangles is less important, with the exception of the anterior, posterior, and submandibular triangles, which have clinical correlates in the current classification of neck disease.

TRIANGLES OF THE NECK

Traditional nomenclature of the surface anatomy of the neck separates it into the anterior and posterior triangles (**Fig. 50–1**). The delineating structure is the sternocleidomastoid muscle, which runs obliquely from its paramedian attachment to the sternum and clavicle posterosuperiorly to the mastoid tip. The anterior triangle is then limited to the midline of the neck medially and

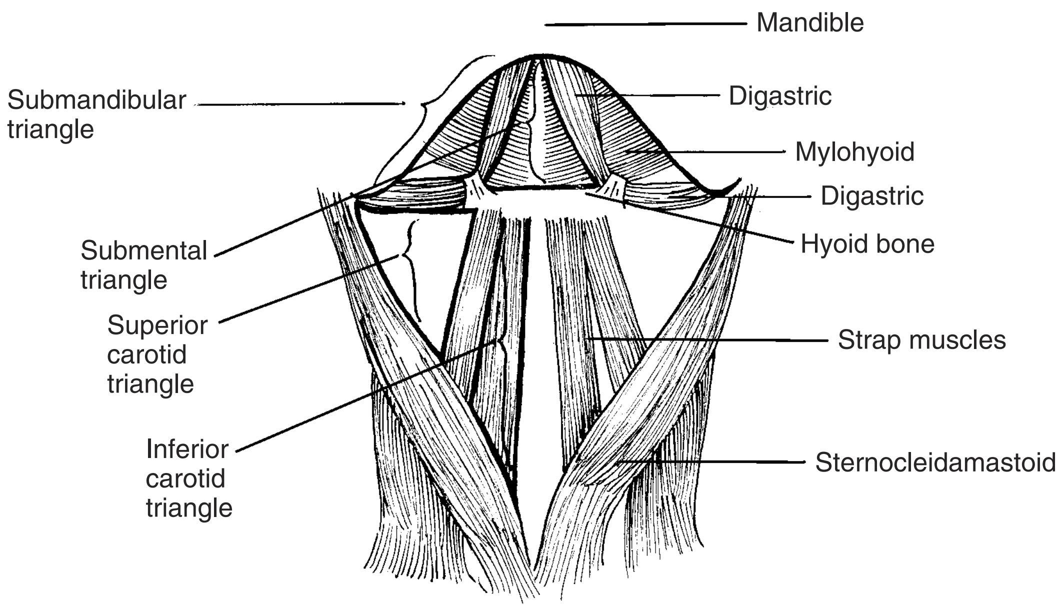


Figure 50–1 *Anterior and posterior triangles of the neck.*

the inferior border of the mandible superiorly. In contrast, the posterior triangle begins at the sternocleidomastoid muscle and is limited posteriorly by the trapezius muscle and inferiorly by the clavicle. The anterior triangle is then further classified as described following here.

Submental Triangle

The submental triangle is bounded by the anterior belly of the digastric muscles laterally and in the hyoid bone inferiorly. It has a triangular shape, with the apex at the mental protuberance and its base at the hyoid bony inferiorly. This triangle, as a midline structure, serves as a lymphatic drainage basin for the lower lip, floor of mouth, and mentum. It holds minimal clinical significance, and in practice this is grouped with the submandibular triangle contents into lymphatic nodal level I.

Submandibular Triangle

The submandibular triangle is bounded by the lower border of the mandible superiorly and by the anterior and posterior bellies of the digastric muscle inferiorly. The contents of this triangle include the submandibular gland, the facial artery and vein, the marginal mandibular nerve, and lymphatic tissue. Additionally, deep within this triangle lie the lingual and hypoglossal nerves (**Fig. 50–2**). The lymphatic tissue is composed of lymph nodes, which are located superior to the submandibular gland, in close proximity to the facial artery and vein, and the marginal mandibular nerve. Occasionally, lymph nodes may be found posterior to the submandibular gland, between it and the parotid gland in the upper neck.

Carotid Triangle

The carotid triangle is a superior component of the anterior triangle of the neck. It is bounded by the posterior belly of the digastric muscle superiorly, by the sternocleidomastoid muscle posteriorly, and by the superior belly of the omohyoid inferiorly. Within it can be found the carotid sheath and its contents, cranial nerves, and lymphatic tissue. Although this area has structures of great clinical import, the classification as "carotid triangle" has no clinical use. Clinical analysis of this area includes the structures within the muscular triangle and all remaining tissue in the anterior triangle inferior to the omohyoid muscle and anterior to the clavicle. This would include the strap muscles and great vessels of the neck.

Posterior Triangle

The posterior triangle is bounded anteriorly by the posterior aspect of the sternocleidomastoid muscle, inferiorly by the clavicle, and posteriorly by the anterior border of the trapezius muscle. The contents of this triangle include the spinal accessory nerve, lymph nodes, and fibrofatty tissue. Deep within this triangle lie the splenius capitis muscle, the levator scapulae muscle, the scalene muscles, and the brachial plexus. However, these all lie underneath the deep layer of the deep cervical fascia and should not be encountered during the course of the neck dissection, with the exception

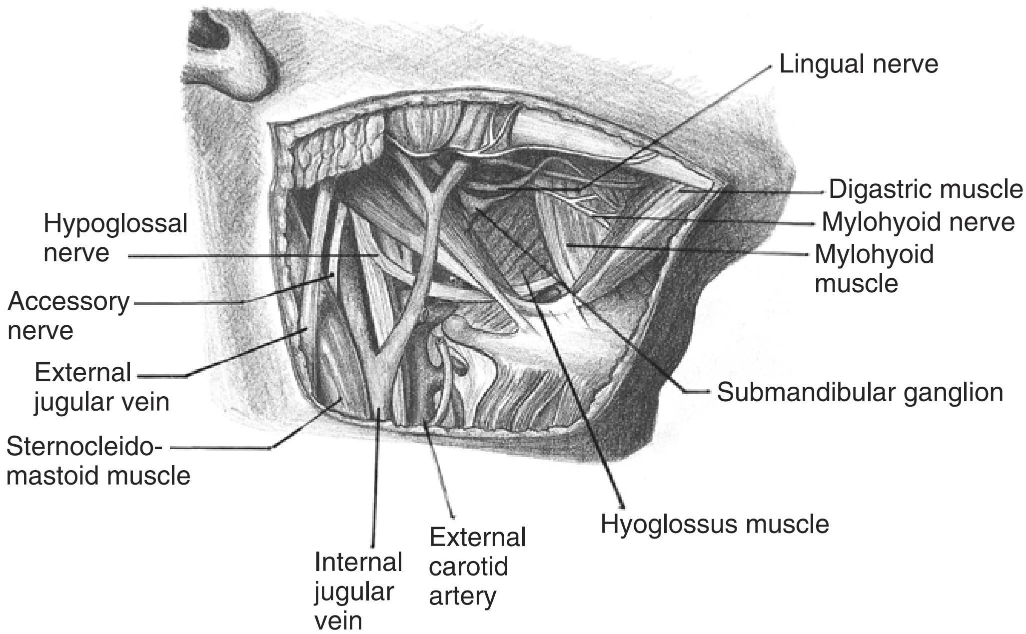


Figure 50–2 *Submandibular triangle and contents (V and XII).*

of visualization of these structures. Additionally, the transverse cervical artery and vein lie in the inferior aspect of this triangle.

MUSCLES

The muscular anatomy of the neck is relatively straightforward when considering the surgical anatomy of the neck. The platysma muscle circles the anterior neck and is the first muscle encountered following a skin incision. It is invested by the superficial layer of cervical fascia and constitutes the reinforcing layer when closing the cervical skin. This thin, fan-shaped muscle begins at the level of the clavicle and continues superiorly over the mandible to join the superficial fascia of the face.

Deep to the platysma, the sternocleidomastoid muscle divides the neck obliquely, originating at the mastoid tip and extending down to the sternum and clavicle. This muscle serves as the clinical dividing point between levels II, III, IV, and V of the neck, which will be discussed later in the chapter. Deep to the sternocleidomastoid muscle, the omohyoid muscle runs obliquely from the hyoid bone to the lateral aspect of the clavicle superficial to the carotid sheath and deep muscles of the neck, such as the scalenes and levator scapulae. Superior, anterior, and deep to the sternocleidomastoid muscle lies the digastric muscle. This muscle is sling shaped, originating from the mandibular tubercle, passing posteroinferiorly to the hyoid bone, where it is tethered to the hyoid bone by a fascial loop, then passing posterosuperiorly to the mastoid tip. The digastric muscle is clinically significant because it is the dividing point between levels I and II, the submandibular triangle, and the upper aspect of the anterior triangle of the neck.

The remainder of the muscles in the neck are not encountered routinely during neck dissection. However, we will describe their anatomy and anatomical relationships at this time. Centrally lie the strap muscles, the sternohyoid, sternothyroid, and thyrohyoid muscles. These muscles provide elevation of the larynx during swallowing, as well as appropriate positioning of the larynx during certain aspects of vocalization. They are not resected during neck dissections, unless they happen to be involved with tumor. The floor of the neck is formed by the anterior, middle, and lateral scalene muscles, as well as levator scapula and splenius capitis muscles. As with the strap muscles, these are not routinely included in surgery involving the neck, unless clinical evidence of tumor involvement is present.

VASCULAR STRUCTURES

Although the critical vascular structures within the neck are primarily related to the carotid sheath, the superficial venous system is important to understand for surgical access to the neck (**Fig. 50–3**). The anterior jugular and external jugular veins lie immediately beneath the deep surface of the platysma muscle and serve to define the appropriate plane of superficial dissection.

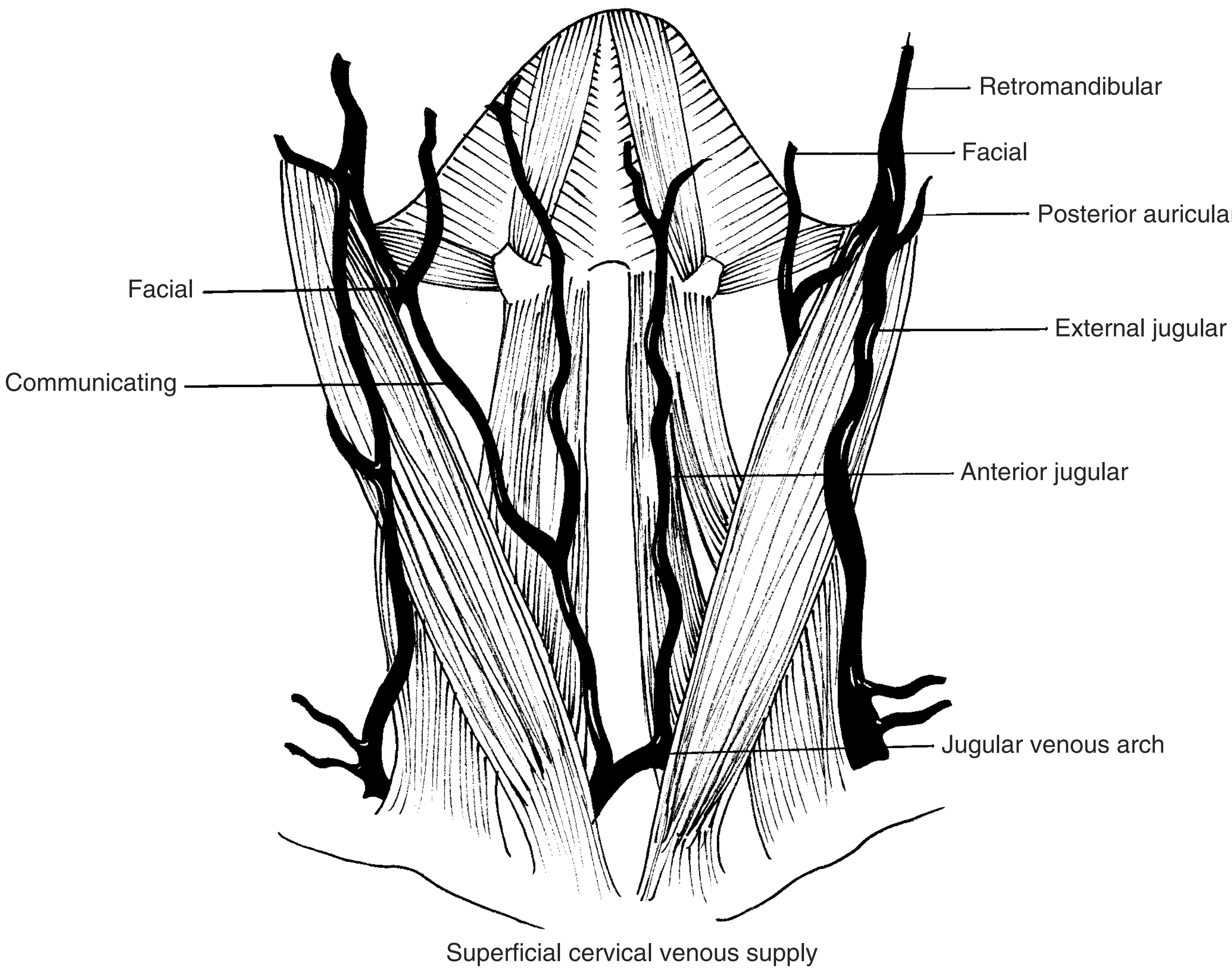


Figure 50–3 *Superficial venous system.*

Therefore, when raising subplatysmal flaps, these veins should be kept intact and on the deep surface of the dissection. Thus the only vascular supply that should be transected during that dissection would be the arterial and venous perforators that pass through the platysma to supply the subdermal plexus of the skin flap. The external jugular vein runs in a craniocaudal direction from the parotid gland superiorly to the clavicle inferiorly. It lies on the superficial surface of the sternocleidomastoid muscle, and is a useful surgical landmark to proper dissection planes. It also crosses the sternocleidomastoid muscle in the area of Erb's point, the point near the midportion of the posterior sternocleidomastoid muscle where the cutaneous sensory nerves become superficial and course anteriorly to supply sensation to the anterior neck. Additionally, in the anterior neck are the anterior jugular and communicating veins. These drain the submental area and continue inferiorly along the strap muscles to connect in the lower neck just superficial to the sternal notch, in the suprasternal space of Burns.

Understanding the vessels associated with the carotid sheath is critical to mastering the surgical anatomy of the neck (**Fig. 50–4**). The internal jugular vein begins its course through the neck cranially in the jugular foramen of the skull and moves inferiorly to join with the subclavian vein to form the brachiocephalic vein. It is often the first great vessel encountered in a dissection of the neck due to its more superficial and lateral positioning. It has many branches emanating from its anterior wall that must be identified and controlled. As the vein exits the skull base, it sends pharyngeal branches medially that have ramifications with other branches off the internal jugular and common facial vein to form a venous plexus overlying the hypoglossal nerve lateral to its emergence from the submandibular space. These venous comitantes can be quite troublesome if inadvertently damaged during dissection of the anterior triangle, and they can lead to

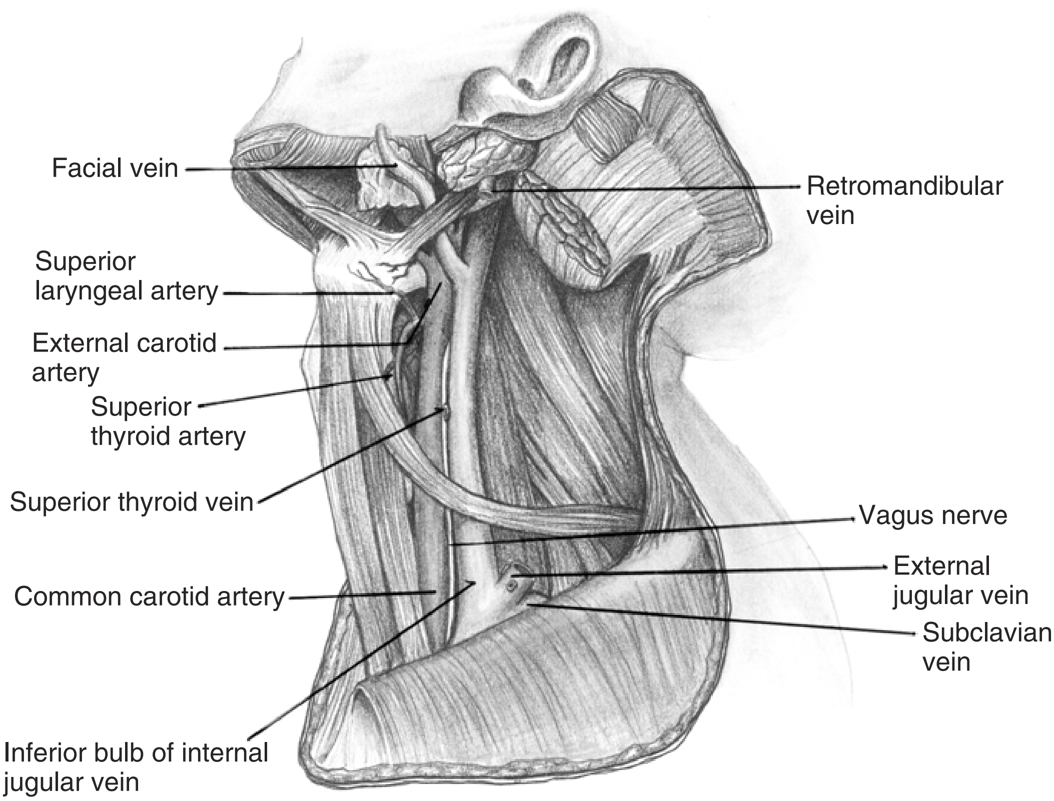


Figure 50–4 *Internal jugular vein and carotid artery.*

damage to the underlying hypoglossal nerve if hasty control of the bleeding is performed. The lingual and superior thyroid veins are next as we progress inferiorly, and the middle thyroid vein is the inferiormost large vessel originating from the internal jugular. As stated above, all major branches off the internal jugular vein originate on the medial aspect of the vein, and this fact should be taken into account during surgery along the jugular vein.

The arterial supply to the neck originates from the carotid artery and its branches. The common carotid, which originates in the root of the neck, courses superiorly deep and medial to the internal jugular vein. The carotid bifurcates at the level of the hyoid bone, sending the branchless internal carotid superiorly with the internal jugular vein to enter the skull base, while the external carotid branches to supply the structures of the neck and face. The carotid bulb is of clinical significance due to the baroreceptors present, which, when stimulated by palpation or dissection, can cause bradycardia in the patient. Branching patterns of the major arteries originating from the carotid are quite variable, and the surgeon should be aware of such variations (**Fig. 50–5**). In general, however, the arteries of clinical significance during; neck dissection; that is, the superior thyroid, facial, lingual, and occipital arteries, are easily identified and fairly constant in their course, if not their origin. The superior thyroid artery is most frequently the first artery off the external carotid, and it lies on the lateral surface of the inferior constrictor and deep to the strap muscles. It should not be divided during routine neck dissection but may be inadvertently damaged during resection of the medialmost contents of levels II and III overlying the thyrohyoid region. The facial artery, in contrast, is often sacrificed during the dissection of the submandibular triangle to facilitate removal of its contents. It is present in the superior submandibular triangle at the facial notch in the mandible and is associated with the perivascular lymph nodes. It enters the submandibular triangle inferiorly deep to the posterior belly of the digastric muscle near the hyoid sling. In this area it is in proximity to the hypoglossal nerve. The lingual artery, which often shares a common trunk with the facial, should not be encountered during routine neck dissection. As is the case for the facial artery in level I, the occipital artery is frequently encountered and divided during the dissection of level II. It crosses the superior portion of the internal jugular vein obliquely to pass underneath the sternocleidomastoid muscle at the mastoid tip and is included in the specimen in this area if necessary. The remainder of the carotid branches are not routinely encountered during

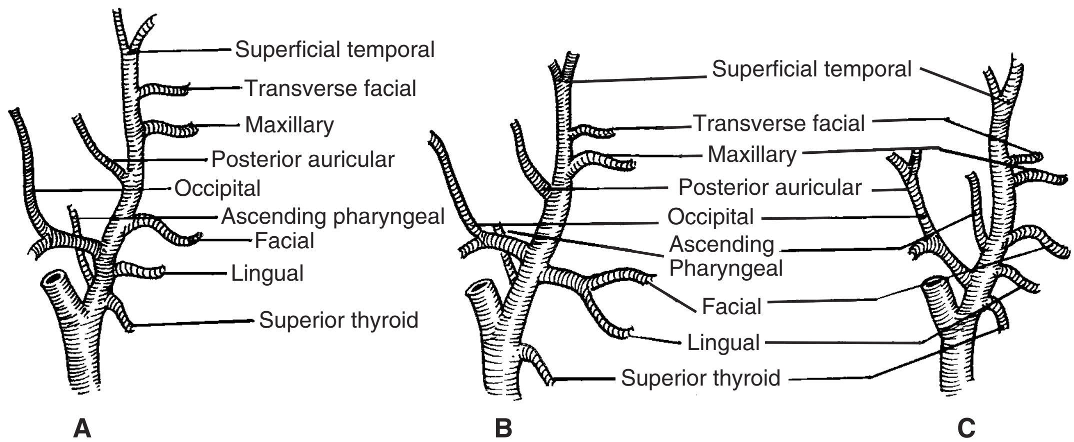


Figure 50–5 *Anomalies of the carotid artery.*

standard neck dissection, and the reader is referred to an anatomy text for further details.

NERVES

Knowledge of the nerves present within the neck, as well as their course, is imperative if one is to operate on the neck. Although there are many nerves, both sensory and motor, that reside in the neck, there are several that are more important clinically. The description that follows will discuss the clinically relevant sensory and motor nerves of the neck, followed by details of the course and anatomical relationships of the lower cranial nerves.

The cervical plexus provides much of the sensory and motor nerve supply to structures of the neck. Contributions from the ventral rami of the second, third, and fourth cervical nerves, with occasional contributions from the first cervical nerve, form this plexus (**Fig. 50–6**). The majority of the sensory nerves enter the more superficial portion of the neck through the posterior triangle, either passing across the sternocleidomastoid muscle from posterior to anterior, or passing through the posterior triangle to supply the posterior cervical and supraclavicular skin through the supraclavicular nerves. These rootlets also provide the motor supply to the deep neck muscles, including the longus capitis, levator scapulae, and middle scalene muscle; there are also contributions from the second cervical nerve to the spinal accessory nerve in some patients. The most clinically relevant of these sensory nerves is the greater auricular nerve, formed from the second and third cervical nerves. This passes around the posterior border of the sternocleidomastoid muscle, emerging at Erb's point, and courses superiorly to supply sensation to the parotid area and the auricle. If damaged or resected during neck dissection, the patient will have permanent anesthesia of the ear

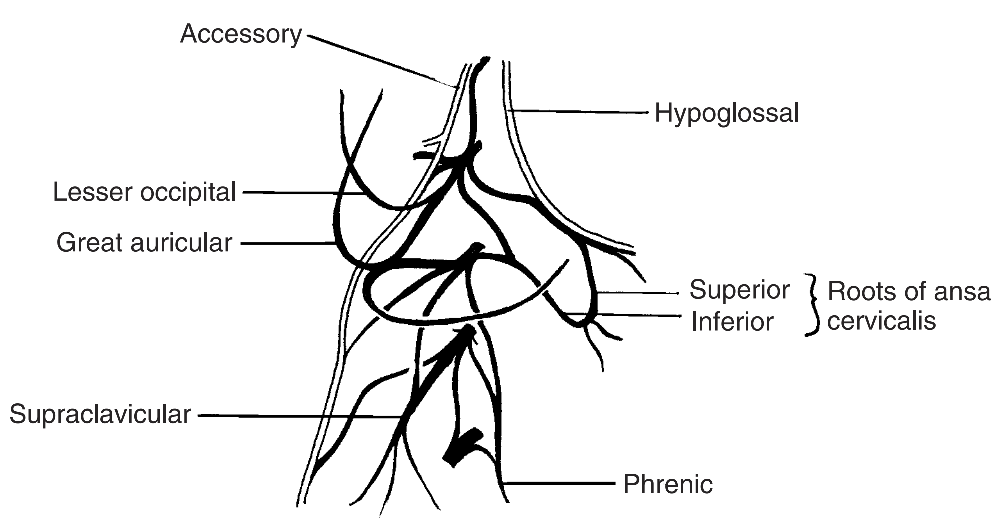


Figure 50–6 *Cervical plexus.*

lobule, which frequently bothers the patient if not forewarned, and may promote frostbite to the area if special attention is not paid to avoiding frostbite in colder climates. As was the case with the superficial veins of the neck, these cutaneous nerves serve as an excellent reminder of the appropriate plane when raising subplatysmal flaps, and they should not be divided, except in their distal branches, during flap elevation.

Unlike the sensory nerves, which exit the deep neck posterior to the sternocleidomastoid muscle, the motor nerves course anteriorly, through the anterior triangle, to reach their muscular destination. The second and third cervical nerves combine to form the ansa cervicalis, which provides motor control to the adjacent strap muscles. These nerves frequently course along the internal jugular vein, on its superficial surface, and may be confused for the vagus nerve by surgeons with limited experience. These nerves are resected during radical and modified radical neck dissections. Their clinical significance lies primarily in the use of innervated strap muscle flaps for laryngeal reconstruction because unilateral sacrifice has no deleterious effect on swallowing. The most important deep contribution of the cervical nerves is the phrenic nerve, which has contributions from the third, fourth, and fifth cervical nerves. The phrenic nerve lies deep to the deep layer of the deep cervical fascia and should not be disturbed during a neck dissection. However, sharp dissection of the medial aspect of the lower neck may transgress this fascia and put the nerve at risk. Therefore, it is imperative to identify this nerve when operating in the medial aspect of level IV in the lower neck. This nerve is unusual in that it originates laterally in the paramedian position and travels medially as it descends into the mediastinum, unlike most nerves in the neck, which travel laterally as they descend, such as the spinal accessory and brachial plexus. The brachial plexus is the last clinically significant contribution of the cervical nerves as the lower neck is dissected. They should never be disturbed because they too lie beneath the deep cervical fascia as they emerge from between the scalene muscles to descend to the upper extremities.

The lower cranial nerves are also critical structures within the neck and are routinely encountered during neck dissection. They are present to some degree within all levels of the neck; therefore, detailed knowledge of their anatomy is of paramount importance. Throughout the course of a neck dissection, branches of the trigeminal, facial, vagus, spinal accessory, and hypoglossal nerves may be encountered. Although the ninth cranial nerve does traverse the superior portion of the neck on its way to the parapharyngeal space, it is rarely visualized during standard neck dissections and will not be discussed further.

All of these nerves are present within the superior neck, and it is in the dissection of the upper neck that they are most frequently encountered. The motor nerve most frequently identified, and injured, in this area is the marginal mandibular nerve. It has several ramifications in the submandibular area that are of note. The cervical branch of the nerve generally takes off the main portion of the marginal nerve prior to the posterior aspect of the submandibular gland, and therefore is frequently divided during elevation of the platysma/cutaneous flap, producing no issue of clinical significance. However, if the cervical branch is observed, it may be traced proximally to identify the main portion of the nerve. The marginal nerve may then be traced distally to ensure its integrity. Many times this is not the case, and the marginal nerve should be sought along the superior aspect of the submandibular gland, where it runs along the lower border of the mandible, superficial to the facial artery and vein. In many instances the nerve may be found immediately inferior to the facial notch, or groove, an indentation in the inferior border of the mandible where the facial vessels emerge from the deep neck to course superiorly in the superficial tissues of the face. Reflection of the facial vessels in the submandibular triangle is often used to assist in protection of the nerve because the nerve will be carried farther superiorly, out of the field of dissection. Extreme care should always be taken when dissecting the submandibular gland, however because the marginal nerve may be present anywhere along the superficial aspect of the gland and has been reported to be found up to 3 cm below the mandible. In addition, branches of the nerve supplying the mentalis muscle commonly course along the lower portions of the submandibular gland in its overlying fascia. These should be preserved if feasible and reflected superiorly with the skin flap. Although they have limited clinical significance, injury may cause asymmetry in the lower lip and mentum with facial expression.

The fifth and twelfth cranial nerves are also present in the submandibular space. The lingual nerve, a branch of the mandibular nerve (the third branch of the trigeminal), courses through the superior aspect of the submandibular triangle and is the only clinically significant branch of the trigeminal nerve encountered during routine neck dissection. It is observed in the uppermost aspects of level I following inferior mobilization of the submandibular gland. It is tethered to the gland by fibers of the parasympathetic submandibular ganglion, which is attached to both the nerve and the gland in this region. Careful dissection is required to avoid inadvertent transection of the lingual nerve in this area. The large caliber of the nerve, as well as its slinglike appearance prior to dividing the submandibular ganglion, should make identification easy. Once the submandibular ganglion is divided, however, the nerve retracts deep into the submandibular space and may be difficult to identify. The hypoglossal nerve is found deep within the submandibular space, deep to the fascia surrounding the gland. It courses anterosuperior from its entry into the submandibular space posterior to the attachment of the digastric muscle to the hyoid bone toward the anterior tongue. It lies on the hyoglossus muscle, whose fibers run perpendicular to the mylohyoid muscle, and it is surrounded by delicate veins that are easily injured if dissected. This nerve lies below the plane of dissection in this region, and it should not be significantly exposed. The hypoglossal nerve may also be identified within Lesser's triangle without removing the submandibular gland. Lesser's triangle is formed by the inferior border of the submandibular gland, following the gland's superior traction, and the superior edges of the posterior and anterior bellies of the digastric muscle.

The hypoglossal and spinal accessory nerves are clinically important within the upper aspect of the neck above the level of the hyoid bone. The location of these nerves is depicted in **Fig. 50–7**. The hypoglossal, as it exits the skull base, courses along the anterior aspects of the internal jugular vein and internal carotid artery to descend to just above the level of the hyoid bone as it courses deep to the digastric and stylohyoid muscles to enter the submandibular triangle. During its descent, it passes lateral to the internal and external carotid arteries, giving off the superior root of the ansa cervicalis. It is tethered into this lower position in the neck by the sternocleidomastoid branch of the occipital artery, around which it hooks to course anteriorly and superiorly. This lower portion of the nerve is covered by a dense venous plexus, the vena comitans, which can provide significant bleeding if inadvertently damaged during dissection. If bleeding is encountered in this area, extreme care must be taken when controlling it to avoid clamping the hypoglossal nerve. Frequently, the nerve is identified anteriorly, just behind the digastric muscle at the level of the hyoid, and traced proximally to free the fibrofatty tissue in the area from the nerve during its dissection.

The spinal accessory nerve, in contrast, is frequently identified superiorly and traced distally into the lower neck. The spinal accessory nerve most frequently courses lateral to the internal jugular vein in the upper neck as it exits the jugular foramen and runs posteroinferiorly. However, up to one quarter of the time it may run behind and deep to the jugular vein in this area, and may even pass through the vein. It then moves posteriorly as it descends, giving off a branch to the sternocleidomastoid muscle. Although the exact relationship between the nerve and the muscle is variable, it most frequently splits the muscle, coursing through it, rather than running on its deep surface. It then runs obliquely through the posterior triangle, entering the triangle ~1 cm above Erb's point. It can be very superficial in the posterior triangle and may be injured or transected during posterior flap elevation if the plane of elevation is too

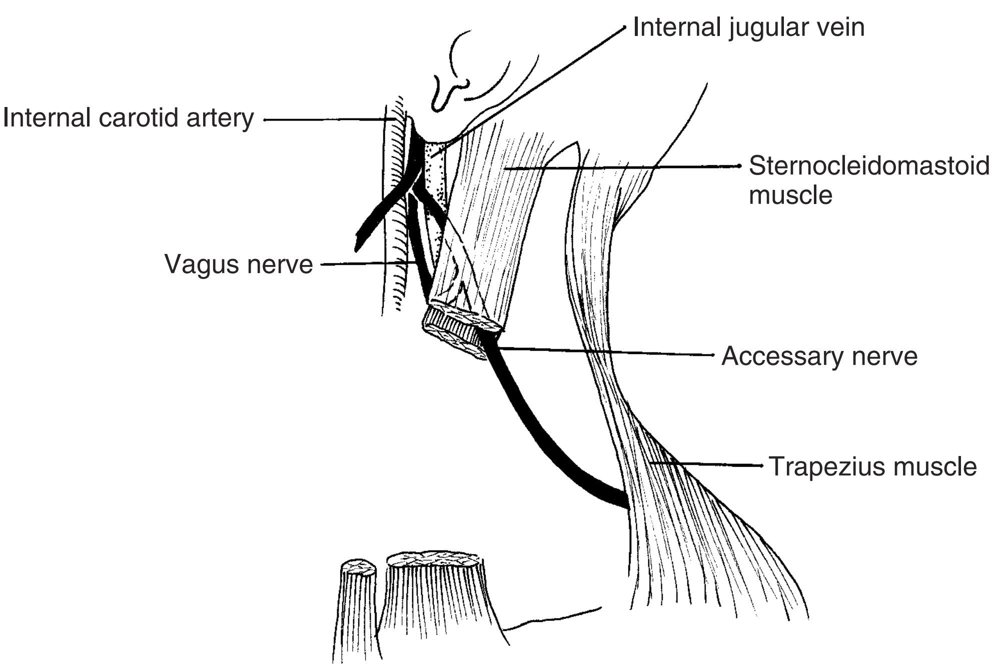


Figure 50–7 *Level II cranial nerves.*

deep. In this area it is just deep to the superficial layer of the deep cervical fascia. It then enters the lower third of the trapezius muscle after passing through the posterior triangle. The nerve is frequently joined, or crossed, by a branch from the second cervical nerve in the area of the sternocleidomastoid muscle, and proximal dissection of the nerve in this area may lead the inexperienced surgeon to follow the cervical nerve deeper into the neck rather than the spinal accessory as it moves superiorly.

Although the vagus nerve runs through many levels of the neck, its anatomical relationships are fairly constant. The nerve originates in the jugular foramen and remains in a position along the lateral aspect of the carotid artery, deep to the medial aspect of the internal jugular vein, as it moves from the skull base to the root of the neck. During its descent, it gives off a pharyngeal branch shortly after exiting the skull base that controls motor function of the upper pharynx. It then gives off the superior laryngeal nerve, which descends deep to the carotid artery to emerge from its medial surface to supply sensory and motor function to the larynx. Farther inferiorly, usually below the clavicles, the recurrent laryngeal nerves take off from the vagus to supply motor and sensory innervation to the larynx as well. With the exception of identification and isolation of the nerve along the carotid sheath, this nerve should not be dissected during a routine neck dissection.

VISCERA

The visceral compartment of the neck is a midline structure containing the trachea, hypopharynx, esophagus, and thyroid gland. A detailed description of the cervical viscera is beyond the scope of this chapter and is presented in the appropriate chapters dealing with those organs.

LYMPHATICS (LEVELS, LANDMARKS, AND DRAINAGE PATTERNS)

Lymphatics within the neck are critical to the spread and management of neoplasms within the head and neck. Knowledge of their location, drainage patterns, and classification are of paramount importance because as there have been reports of as many as 300 lymph nodes present in the head and neck, and such information allows appropriate dissection of the cervical tissues. There have been many classification schemes proposed to describe lymphatic disease within the neck. Due to the vagaries and complexities of many of these schemes, an effort was undertaken in the early 1990s to standardize the terminology describing the locations of the cervical lymphatics. The resultant terminology has been adapted by the major head and neck oncology societies and forms the basis of universal descriptions of neck disease. The older method of referring to the nodal basins as upper, middle, and lower jugular and jugulodigastric has been replaced by the numbering system detailed following here and illustrated in **Fig. 50–8**.

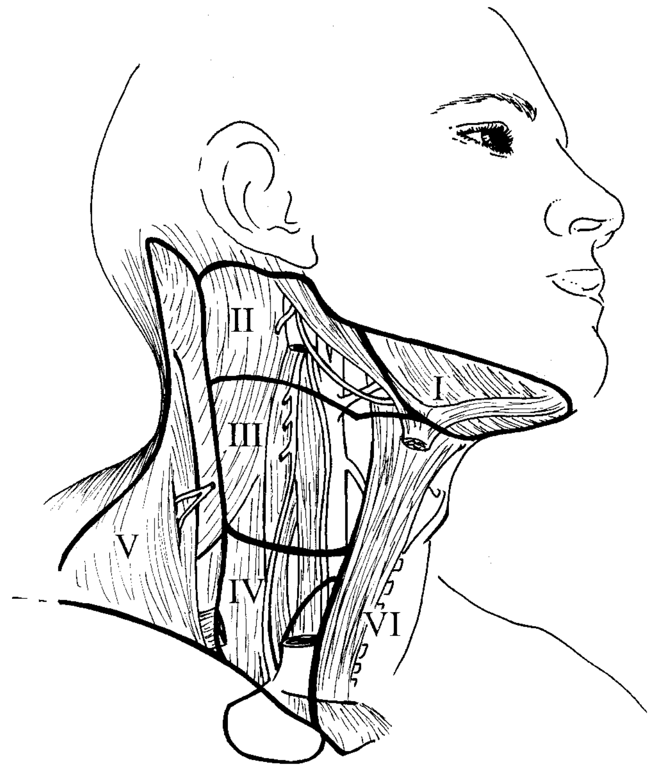


Figure 50–8 *Nodal levels.*

Level I

This nodal basin is formed by the submental and submandibular nodal groups present in prior classification schemes. It is bounded superiorly by the lower border of the mandible, inferiorly by the posterior belly of the digastric muscle and the hyoid bone, posteriorly by the parotid gland, and anteriorly by the midline. The submandibular gland is included in this specimen when the nodes are removed. The lymph nodes in this level are primarily located along the superior aspect of the submandibular gland, near the border of the mandible. They are in close proximity to the facial vessels and are commonly referred to as the pre- or postvascular lymph nodes. Additionally, nodes may be found along the posterior aspect of the gland, as well as within the submental area of level I.

There is a predictable drainage pattern to the lymph nodes present in level I. There are several sites within the head and neck whose primary echelon of lymph nodes may be level I. These sites include the lips (upper and lower), the tongue and floor of the mouth, the sublingual and submandibular glands, alveolus, buccal mucosa, hard palate, and the anterior nose.

Level II

Level II lies in close proximity to the carotid sheath structures and is the current classification for what was previously referred to as the upper jugular nodal group. It, like all nodal levels within the anterior triangle, is bounded anteriorly by the lateral border of the sternohyoid muscle and posteriorly by the posterior border of the sternocleidomastoid muscle. Clinically, during the course of a selective neck dissection, the posterior border of the sternocleidomastoid muscle may not be readily apparent due to distortion from retraction. Therefore, in the case of a selective neck dissection, the posterior cervical rootlets are used to define the posterior border of levels II, III, and IV. The superior border of this level is the skull base, and the inferior demarcation is the carotid bifurcation (surgical landmark), which correlates clinically with the hyoid bone (clinical landmark). Other, nonlymphatic, structures found within this level include the carotid artery, internal jugular vein, vagus nerve, hypoglossal nerve, and spinal accessory nerve. Additionally, during the course of dissection, one will encounter the sternocleidomastoid branch of the occipital artery, the structure that is felt to be responsible for the descent of the hypoglossal nerve into the neck during development. The main portion of the occipital artery will course transversely across the superior aspect of level II and is encountered during dissection of this area.

As was the case for level I, there are predictable patterns of spread to level II as well. Level II has the highest number of potential primary sites. This area may be a primary drainage basin of the paranasal sinuses, parotid gland, oropharynx (soft palate, faucial arch, tonsils, and lateral pharyngeal walls), oral cavity (alveolar ridge, tongue, floor of mouth), nasopharynx, tongue base, supraglottic larynx, hypopharynx, or cheek skin.

Level III

Level III lies immediately inferior to level II in the neck and shares some of the same boundaries. As in level II, the anterior and posterior boundaries of this level are the sternohyoid and sternocleidomastoid muscles, respectively. The surgical definition remains that described in level II; namely, the cervical rootlets. Superiorly, the level is delineated by the carotid bifurcation, or hyoid bone clinically, and the inferior border is the upper aspect of the omohyoid muscle (surgical landmark) or the cricothyroid notch (clinical landmark). This area corresponds to what was previously referred to as the middle jugular nodal group. As was true for level II, the nonlymphatic structures present include the carotid artery, internal jugular vein, superior thyroid vessels, and vagus nerve.

Isolated adenopathy in this area is suggestive of several potential primary sites. These include many of those sites that may also present in level II. These sites include the tongue base, supraglottic larynx, glottic larynx, subglottic larynx, hypopharynx, nasopharynx, tongue, floor of mouth, and thyroid gland.

Level IV

Level IV is the inferior continuation of the tissue surrounding the carotid sheath and was referred to as the lower jugular group in the past. It is bounded by the sternohyoid and sternocleidomastoid muscles in the same fashion as levels II and III and starts at the omohyoid muscle superiorly and continues inferiorly to the level of the clavicle. Along with the fibrofatty and lymphatic contents of this area course the carotid sheath structures present in levels II and III. Additional structures include the middle thyroid vein, the thoracic and lymphatic ducts from the chest, the medial aspect of the subclavian vein, and occasionally the transverse cervical vessels. Although the phrenic nerve is not technically within the contents of this level, because it lies beneath the deep layer of the deep cervical fascia, it does lie in the floor of this space; therefore, it must be identified when dissecting this aspect of the neck. Unlike the other nerves running in the floor of the neck in level IV, the phrenic nerve courses lateral to medial as it descends into the mediastinum. Nerves contributing to the brachial plexus run medial to lateral as they descend through the neck, although they are primarily visualized in the course of dissecting level V.

Given the inferior location of level IV within the neck, fewer tumor sites present primarily as isolated level IV metastases. These sites may include the glottic or subglottic larynx, the hypopharynx, the cervical esophagus, and the thyroid gland. Isolated cases of level IV being the primary echelon drainage for the tongue have also been reported.

Level V

Level V occupies the entire posterior triangle, from the mastoid tip and skull superiorly to the clavicle inferiorly. It is bounded by the posterior border of the sternocleidomastoid muscle anteriorly and the anterior border of the trapezius muscle posteriorly. The commonly referred to supraclavicular nodes lie within this level as well. Few structures run within this level, as it is primarily composed of fibrofatty and lymphatic tissue. However, those structures coursing through this region include the spinal accessory and the supraclavicular nerves, as well as the transverse cervical vessels and the posterior belly of the omohyoid muscle. Primary, isolated metastases to level V are very uncommon and usually herald primary disease in unusual areas of the head and neck, or outside the head and neck. Superior masses may originate from the scalp or postauricular area, and inferior masses may originate from the thyroid gland, hypopharynx, cervical esophagus, or infraclavicular organs. Abdominal, thoracic, or breast malignancy must always be considered in the evaluation of masses present within the inferior aspect of level V.

Level VI

Level VI is a midline region of the neck that is not included in standard neck dissections. It is composed of tissues within the anterior compartment of the neck surrounding the central, midline visceral structures extending from the hyoid bone superiorly to the sternal notch inferiorly, and is bounded laterally by the medial aspect of the carotid sheaths. Contained within this level are lymph nodes within the perithyroidal, pretracheal, paratracheal, precricoid, and paraesophageal nodal subgroups. The strap muscles, thyroid and parathyroid glands, and laryngeal structures lie within this level. Lymph nodes in this level are involved in the spread of disease from the thyroid gland, larynx, hypopharynx, and cervical esophagus.

NECK DISSECTIONS

En bloc resection of cervical lymph nodes and associated structures is often attributed to Crile in the early 1900s, although earlier reports of similar resections have been identified. The radical neck dissection, as we know it, was popularized by Hayes Martin in the 1950s and had remained largely unchanged until reports of successful modifications were published in the 1970s by surgeons such as Bocca, who popularized the “functional” neck dissection, and in the 1980s by Byers, who promulgated the concept of selective neck dissection. Because all neck dissections are based on radical neck dissection, the discussion will proceed from the radical neck dissection to those dissections of lesser magnitude.

A radical neck dissection provides complete removal of all fibrofatty tissues and lymph nodes within the neck from levels I through V. This extends from the inferior border of the mandible superiorly, to the clavicle inferiorly, and the lateral aspect of the sternohyoid muscle anteriorly (with the exception of level I) to the anterior aspect of the trapezius muscle posteriorly (**Fig. 50–9**). In addition, the spinal accessory nerve, sternocleidomastoid muscle, and internal jugular vein are all removed.

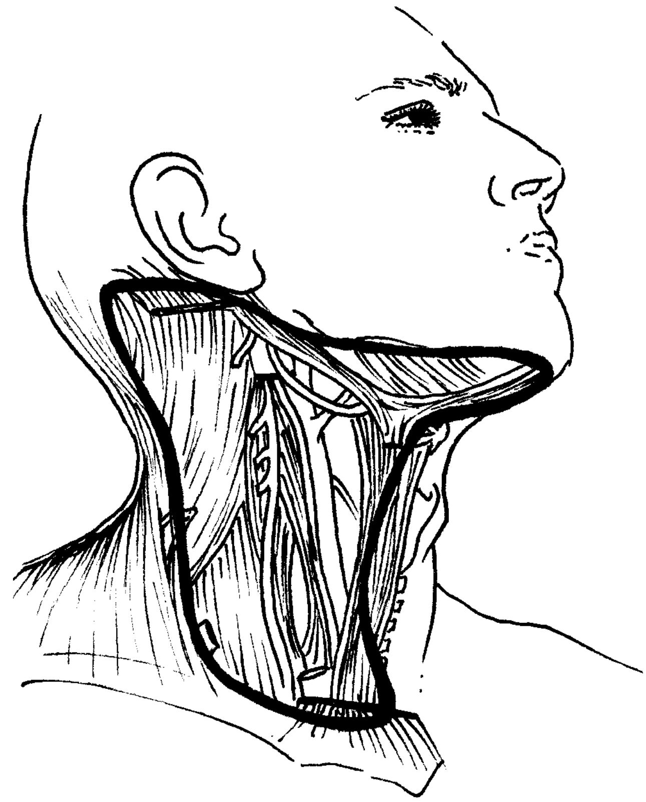


Figure 50–9 *Radical neck dissection.*

This provides the most complete standard dissection of the neck. The carotid artery, or major branches, hypoglossal nerve, vagus nerve, and visceral structures are not included in the dissection. Inclusion of any of these structures defines an extended radical neck dissection and is performed if the disease burden dictates removal of these structures. The radical neck dissection is currently most commonly performed in the presence of advanced neck disease, N3. (See **Table 50–1** for definition of nodal stages.)

Table 50–1 1997 AJCC NODAL CLASSIFICATION SYSTEM

Node	Definition
N0	No palpable regional lymph nodes
N1	Involvement of a single ipsilateral lymph node 3 cm or less in maximal dimension
N2a	Involvement of a single ipsilateral lymph node more than 3 cm but less than 6 cm maximal dimension
N2b	Involvement of multiple ipsilateral lymph nodes all less than 6 cm maximal dimension
N2c	Involvement of bilateral or contralateral lymph nodes all less than 6 cm maximal dimension
N3	Involvement of any lymph node, regardless of side, greater than 6 cm in maximal dimension

AJCC, TK

Less complete neck dissections include the modified radical neck dissection and a variety of selective neck dissections. Modified radical neck dissection has been used extensively in the management of head and neck cancer. A modified radical neck dissection is defined as the same lymphatic resection performed by a radical neck dissection (levels I–V) but with preservation of one of the nonlymphatic structures sacrificed during a radical neck dissection. Modifications may include sparing any one or a combination of the spinal accessory nerve, internal jugular vein, or sternocleidomastoid muscle. The most common modification is sparing the spinal accessory nerve to minimize shoulder morbidity following neck dissection. Modified radical neck dissections are usually employed for early and intermediate stage neck disease (N1 and N2) or following chemotherapy or radiotherapy.

Recent reports have demonstrated that all lymphatic structures do not have to be removed in all cases of head and neck cancer resection. The practice of subtotal lymphatic resection has resulted in selective neck dissections. In the course of these operations, select lymphatic groups are removed while all nonlymphatic structures are preserved, including the internal jugular vein, sternocleidomastoid muscle, and spinal accessory nerve. Any combination of lymphatic levels may be included in the dissection, although at least three contiguous levels are most often removed during these procedures. The most common include the supraomohyoid neck dissection, which removes levels I, II, and III; the lateral neck dissection (levels II, III, and IV); and the posterolateral neck dissection (levels II, III, IV, and V). In general, the supraomohyoid dissection is used for oral cavity lesions, the lateral neck dissection is used for oropharyngeal and laryngeal tumors, and the posterolateral dissection is used for certain laryngeal and hypopharyngeal lesions as well as auricular lesions such as melanoma. These limited neck dissections are most often performed "electively" when nodal metastases are not clinically apparent, although some advocate their therapeutic efficacy in early (N1) nodal disease.

NECK STAGING

The staging of cervical malignant adenopathy is consistent among most head and neck primary sites. These criteria change intermittently as the staging criteria are revised to provide more accurate prognostic information. Nodal spread of cancer is a poor prognostic factor by itself, reducing survival by up to half, stage for stage, but prognosis is also worse as the nodal stage increases. The current nodal classification system is listed in **Table 50–1**.

SUMMARY

The anatomy of the neck continues to fascinate and enthrall medical students and physicians. Mastery of the normal anatomical course of cervical structures, as well as their three-dimensional relationships to one another, is critical to performing successful surgery in the neck. These anatomical factors form the basis of the current surgical techniques used to manage cancer of the head and neck, particularly in the presence of cervical metastases.

Suggested Readings

Hollinshead WH. Anatomy for Surgeons. 3rd ed. New York: JB Lippincott; 1982

Robbins KT, Medina JE, Wolfe GT, et al. Standardizing neck dissection terminology. Arch Otolaryngol Head Neck Surg 1991;117:601–605

Putz R, Pabst R, Taylor AN, eds. Atlas of Human Anatomy. 12th ed. Philadelphia: Williams & Wilkins; 1997

Self-Test Questions

For each question answer true or false for the following statements. To check your answers, see Answers to Self-Tests on page 717.

1. The nerves that are most frequently encountered during neck dissection include the facial, glossopharyngeal, vagus, spinal accessory, and hypoglossal nerves.
2. The relationships between the branches of the external carotid artery are consistent in all patients.
3. Current classification schemes of the cervical lymphatics have both a clinical and an anatomical localization, allowing uniformity in the reporting of nodal locations.
4. There are consistent landmarks that may be used to localize cranial nerves during surgery on the neck.

Chapter 51

SURGICAL ANATOMY OF THE SKULL BASE AND CRANIAL NERVES

Joseph Feghali and Dorothy Frenz

Otolaryngologists, by the nature and location of the organs that they treat, develop a thorough understanding of and familiarity with the structures of the head and neck. Of these structures, the cranial nerves and skull base foramina deserve particular attention. This chapter will review various aspects of the anatomy and physiology of the cranial nerves as they relate to the fields of otolaryngology, neurotology, and skull base surgery.

THE OLFACTORY NERVE (CN I)

The olfactory nerve [cranial nerve (CN) I] is a special sensory nerve that carries the sensation of smell or olfaction. The olfactory system consists of the olfactory nerve and other associated structures.

The olfactory epithelium is usually located in the roof of the nasal cavity, the superior portion of the nasal septum, and the epithelial lining of the superior nasal turbinate. The olfactory epithelium is kept moist by the secretions of the olfactory glands. These secretions dissolve the inhaled aromatic scents.

The neurosensory cells are the primary sensory neurons. Their cell bodies reside within the olfactory epithelium below the level of the cribriform plate. They have peripheral and central processes. The peripheral processes are bathed within the secretions of the olfactory glands and act as sensory receptors. The central processes converge into multiple bundles that traverse the cribriform plate and synapse with the secondary sensory neurons of the olfactory bulb.

The olfactory bulb is a rostral enlargement of the olfactory tract. The olfactory tract travels on the undersurface of the frontal lobe. The olfactory tract and bulb are commonly referred to as the olfactory nerve even though they are mostly composed of secondary sensory neurons and axons. The various cells of the olfactory bulb project through postsynaptic fibers to a variety of central olfactory areas. These, in turn, communicate with autonomic centers for visceral responses such as salivation in response to pleasant smells of food, or nausea in response to unpleasant odors.

Clinical Considerations for Otolaryngologists

- Nasal allergies and polyps can cause enough nasal obstruction to cause anosmia or hyposmia.
- Viral infections can cause damage to the olfactory sensory epithelium and result in anosmia or hyposmia.
- There is recent evidence that some medications (e.g., cisplatin) can cause damage to the olfactory sensory neuroepithelium and result in anosmia or hyposmia.
- Skull fractures can result in a severance of the primary sensory neurons at the level of the cribriform plate, resulting in anosmia.
- Esthesioneuroblastomas are malignant tumors that arise from the olfactory epithelium at the roof of the nose. They usually present with unilateral epistaxis and a nasal mass.

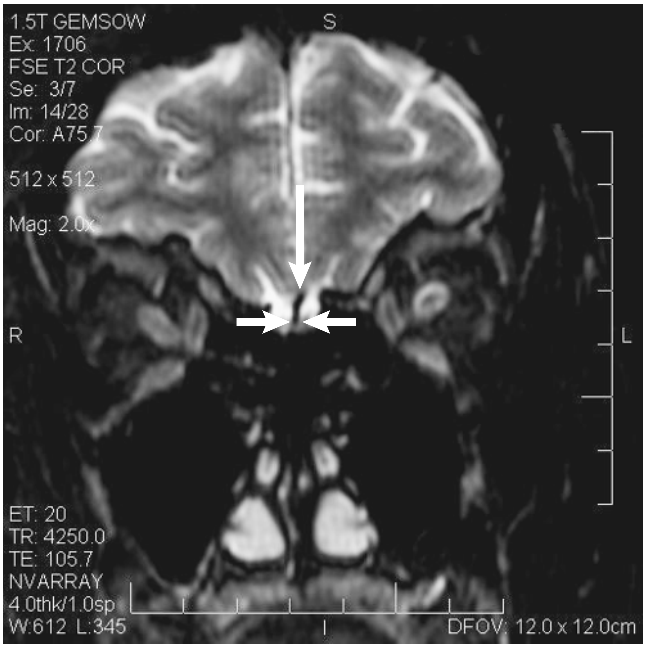


Figure 51–1 *Magnetic resonance imaging (MRI) view of the olfactory nerves (short arrows) on either side of the crista galli. (long arrow).*

- Frontal lobe masses and olfactory groove meningiomas can cause anosmia or hyposmia.
- Olfactory hallucinations may be a symptom of temporal lobe seizures. The associated phantom smell is typically unpleasant (**Fig. 51–1**).

THE OPTIC NERVE (CN II)

The optic nerve (CN II) is a special sensory nerve that transmits visual information from the eye to the visual centers of the brain. Light enters the eye through the pupil and enters the deep layers of the retina. There, light energy is transduced into electric energy by the photoreceptor cells, the rods and cones. The bipolar cells are the primary sensory neurons in the eye. They are also within the retina and receive information from the rods and cones. They then transmit the information to the ganglion cells, which are the secondary sensory neurons. The axons of the ganglion cells converge toward the optic disk, go through the sclera, and exit the eye globe as the optic nerve. The optic nerve then passes through the optic canal. The optic canal is a bony canal within the lesser wing of the sphenoid bone. Like the olfactory nerve, the axons within the optic nerve are secondary sensory cell axons. Thus, technically, the optic nerve is a tract rather than a nerve, but it is typical and customary to call it the optic nerve up to the level of the optic chiasm.

The right and left optic nerves meet and join at the optic chiasm. At the chiasm, approximately half of the fibers from each side cross the midline. The fibers from the nasal half of the retina cross at the chiasm, and the fibers from the temporal half of the retina continue uncrossed within the optic tracts. Most of the fibers within each tract continue on to the lateral geniculate body and eventually terminate at the primary visual cortex in the occipital lobe.

A complete review of the physiology of the eye is beyond the scope of this chapter, but it is important for otolaryngologists to understand and remember the following:

- Because the images on the retina are reversed, the right visual field from both eyes is viewed by the left hemisphere, and the left visual field from both eyes is viewed by the right hemisphere.
- At the chiasm, the nerve fibers that cross the midline carry the visual information from the bitemporal (peripheral) fields of vision.
- Increased intracranial pressure is transmitted to the optic disk as papilledema.
- The optic nerve is really the central nervous system (CNS) tract and can be involved in CNS diseases such as multiple sclerosis.

Figure 51–2 *The optic nerve has been cut on the right to show its relation to the internal carotid artery. The optic nerve (II) is intact on the left and is seen to reach the area of the chiasm. ICA, internal carotid artery; PCA, posterior cerebral artery. (From Sen C, Chen CS, Post KD. Microsurgical Anatomy of the Skull Base and Approaches to the Cavernous Sinus. Stuttgart: Georg Thieme Verlag; 1997:45.)*

Clinical Considerations for Otolaryngologists

- Trauma to the face and orbit can result in injury at the level of the optic canal and foramen. The area of the optic canal can be decompressed through intracranial and transnasal approaches.
- Pituitary gland adenomas compress the optic chiasm and can present with bilateral peripheral visual field cuts.
- Otomastoiditis can result in venous dural sinus thrombosis and increased intracranial pressure and papilledema. In severe cases, increased intracranial pressure can result in visual loss.
- Demyelinating diseases such as multiple sclerosis can cause dizziness and visual symptoms (**Figs. 51–2, 51–3**, and **51–4**).

THE OCULOMOTOR NERVE (CN III)

The oculomotor nerve (CN III) has two components. The first component is a somatic motor (general somatic efferent) that plays an important role in the movement of the eye and the elevation (opening) of the upper eyelid. The second component is visceral motor (general visceral efferent). This second component provides the parasympathetic supply of the constrictor pupillae and the ciliary muscles of the eye.

The general somatic efferent component of the oculomotor nerve plays a major role in the extraocular movement of the eye, but this role is coordinated with that of CN IV and CN VI. Six extraocular muscles control the movement of each eye. These muscles help maintain simultaneous fixation of both eyes on moving objects. This complex function and coordination are mediated through central connections in the cortex and brainstem.

The general somatic efferent component of the oculomotor nerve controls the function of four of the six extraocular muscles. These are the superior rectus,

Figure 51–3 *Left optic nerve (long arrow) and optic chiasm (short arrow) as seen on magnetic resonance imaging.*

medial rectus, inferior rectus, and inferior oblique muscles. The oculomotor nerve also supplies the levator palpebrae superioris muscle.

The somatic motor oculomotor nucleus is situated in the midbrain at the level of the superior colliculus. It supplies the ipsilateral inferior rectus, inferior oblique, and medial rectus muscles. It also supplies the contralateral superior rectus muscles. The oculomotor nucleus supplies the levator palpebrae superioris bilaterally. Fibers from the oculomotor nucleus travel to the area of the interpeduncular fossa at the junction of the midbrain and pons. There they join parasympathetic fibers from the Edinger-Westphal nucleus to form the oculomotor nerve. The oculomotor nerve then travels between the posterior cerebral and superior cerebellar arteries. It pierces the dura and enters the most superior aspect of the cavernous sinus just superior to the trochlear nerve. It then enters the superior orbital fissure through the tendinous ring. Within the orbit, it divides into two motor divisions and the visceral motor division, which enters the ciliary ganglion. The superior motor division innervates the superior rectus and levator palpebrae muscles. The inferior division divides into three branches that innervate the inferior rectus, inferior oblique, and medial rectus muscles. The medial rectus muscle adducts the eye, and the inferior rectus acts in downward gaze. The superior rectus and the inferior oblique act synergistically to elevate the eye.

The visceral motor Edinger-Westphal nucleus is located in the midbrain. Its preganglionic fibers join the oculomotor nerve and run along with it through the cavernous sinus and superior orbital fissure. Once within the orbit they leave the nerve to the inferior oblique muscle and enter the ciliary ganglion near the apex of the cone of the extraocular muscles. The postganglionic fibers enter the globe at the level of the optic nerve and travel to innervate the ciliary and constrictor pupillae muscles. The ciliary muscles control the curvature of the lens. The constrictor pupillae controls the size of the pupil.

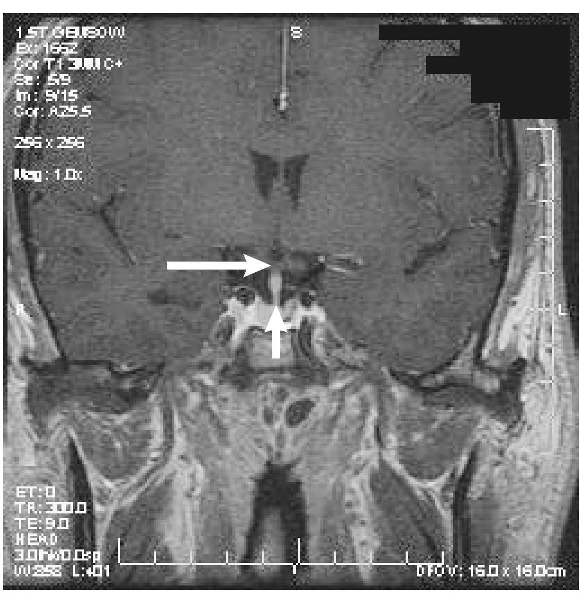


Figure 51–4 *Pituitary stalk (short arrow) in relation to the optic chiasm (long arrow) as seen on magnetic resonance imaging (coronal section).*

Clinical Considerations for Otolaryngologists

- An oculomotor ophthalmoplegia (lower motor neuron) lesion is characterized by strabismus, ptosis, pupillary dilation, and lack of accommodation. The strabismus is characterized by abduction and downward deviation of the eye secondary to the unopposed action of the lateral rectus (CN VI) and superior oblique (CN IV) muscles. The ptosis is due to the paralysis of the levator palpebrae and unopposed action of the orbicularis oculi muscle. The pupillary dilation and lack of accommodation is due to the paralysis of the general visceral efferent branches of the oculomotor nerve.
- Pathological conditions within the cavernous sinus, superior orbital fissure, and orbit can cause oculomotor nerve palsies that are of immediate concern to otolaryngologists. Disseminated infection, inflammation, sinusitis, trauma, and

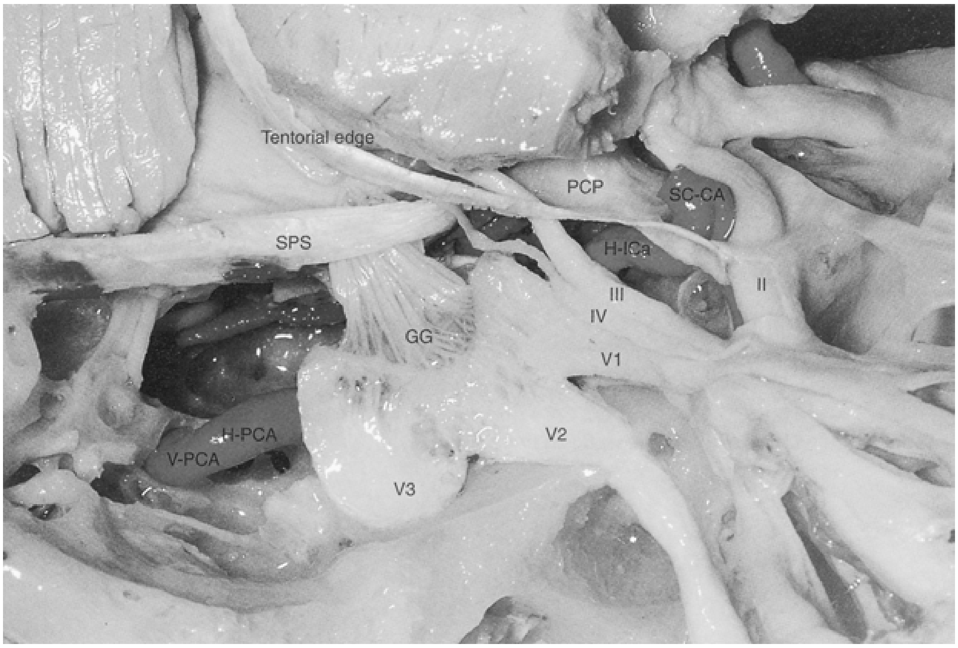


Figure 51–5 *Segments of intracavernous carotid artery. GG, gasserian ganglion; H-ICa, horizontal segment of intracavernous carotid artery; H-PCA, horizontal segment of petrous carotid artery; PCP, posterior clinoid process; SC-CA, supraclinoid carotid artery; SPS, superior petrosal sinus; V-PCA, vertical segment of petrous carotid artery. The oculomotor nerve (III) is shown entering the most superior aspect of the cavenous sinus just superior to the trochlear nerve (IV). (From Sen C, Chen CS, Post KD. Microsurgical Anatomy of the Skull Base and Approaches to the Cavernous Sinus. Stuttgart: Georg Thieme Verlag; 1997:21.)*

similar conditions are usually associated with palsy of other cranial nerves that traverse the cavernous sinus and/or the superior orbital fissure. These conditions have to be differentiated from neurological and neurosurgical conditions such as aneurysms of the posterior cerebral and superior cerebellar arteries or other intracranial pathology (**Fig. 51–5**).

THE TROCHLEAR NERVE (CN IV)

The trochlear nerve (CN IV) is a somatic motor (general somatic efferent) nerve that supplies the superior oblique muscle of the eye.

The superior oblique muscle is one of six extraocular muscles that contribute to the movement of the eye. Its action is coordinated with that of the other five muscles and with the corresponding muscles of the contralateral eye. Coordinated movement of all extraocular muscles helps both eyes maintain simultaneous movement. Central connections help maintain coordination between the movement of all extraocular muscles.

The nucleus for the trochlear nerve is located in the tegmentum of the midbrain at the level of the inferior colliculus. The fibers of the trochlear nerve travel dorsally and cross to the contralateral side, making it that each superior oblique muscle is innervated by the contralateral trochlear nucleus. They then curve around the cerebral peduncle. The fourth nerve then travels along with the third nerve between the posterior cerebral and superior cerebellar arteries and pierces the dura at the angle between the free and the attached borders of the tentorium cerebelli. It then enters the cavernous sinus and runs along the lateral wall of the sinus inferiorly and laterally to the oculomotor nerve. The cranial nerve enters the orbit through the superior orbital fissure above the tendinous ring. It then crosses medially close to the roof of the orbit, running diagonally across the levator palpebrae to reach the superior oblique muscle. When it contracts, the superior oblique muscle causes the eye to rotate inward and the eye globe to move laterally and downward.

Clinical Considerations for Otolaryngologists

- Paralysis of the trochlear nerve results in outward rotation of the eye and weakness of downward gaze. Patients with a fourth nerve palsy complain of difficulty looking down while going down the

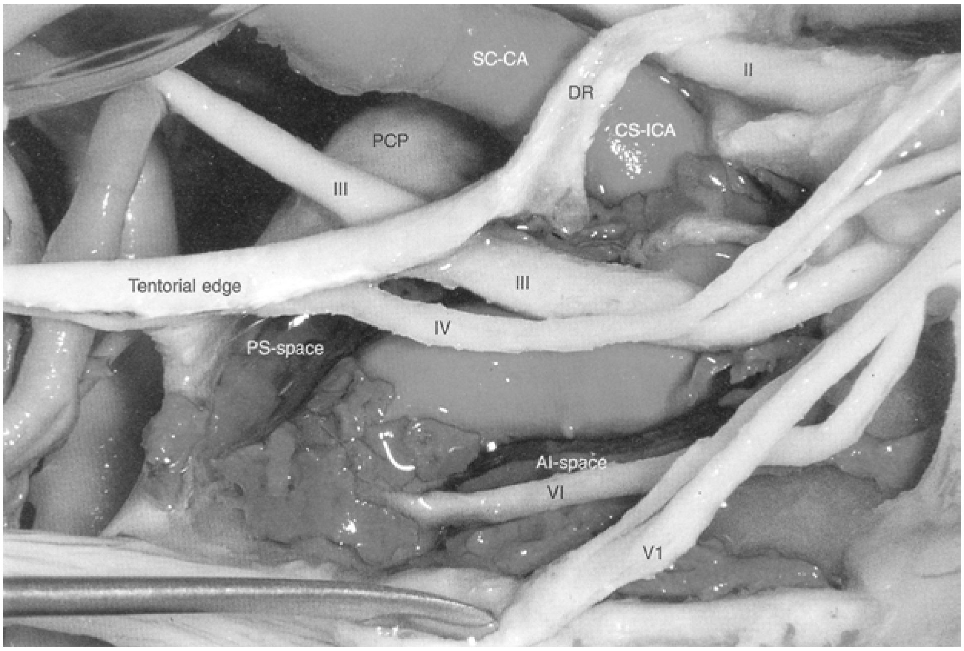


Figure 51–6 *The trochlear nerve (IV) is shown in its close relation to the tentorial edge. This anatomical proximity puts the trochlear nerve at risk during procedures that require severing the tentorial edge. AI-space, anteroinferior venous space; CS-ICA, clinoid segment of internal carotid artery; DR, distal ring; PCP, posterior clinoid process; PS-space, posterosuperior venous space; SC-CA, supraclinoid carotid artery. (From Sen C, Chen CS, Post KD. Microsurgical Anatomy of the Skull Base and Approaches to the Cavernous Sinus. Stuttgart: Georg Thieme Verlag; 1997:66.)*

stairs. They also tend to tilt their head toward the normal side to compensate for the rotation of the eye on the affected side.

- The trochlear nerve has several characteristics that make it susceptible to injury. It is the cranial nerve with the longest intracranial course. It is the smallest cranial nerve with the least number of fibers. Location of the fourth nerve just deep to the tentorium cerebelli puts it at risk when sectioning the tentorium during the subtemporal approach to the midbrain and the extended middle fossa approach to the petroclival area (**Fig. 51–6**).

THE TRIGEMINAL NERVE (CN V)

The trigeminal nerve (CN V) has two components. The first component is a general sensory (general somatic afferent) component that supplies the facial sensations and sensations from multiple areas within the head and neck. The second component is branchial motor (special visceral efferent) and supplies the muscles of mastication and other smaller muscles within the head and neck.

The sensory component connects to the sensory nucleus, which is the largest of the cranial nerve nuclei. The sensory nucleus extends from the midbrain to the second cervical segment (nucleus of the spinal tract). The motor or masticator nucleus is located in the midpons. The trigeminal nerve emerges from the lateral surface of the pons. It has a major sensory component and a smaller sensory component. The nerve then travels to the trigeminal ganglion, which sits in Meckel's cave, a bony depression in the floor of the middle cranial fossa. After exiting Meckel's cave, the nerve divides into three major divisions: the ophthalmic branch (V_1), the maxillary branch (V_2), and the mandibular branch (V_3). These branches exit the skull via the superior orbital fissure, foramen rotundum, and foramen ovale, respectively.

The sensory component of V_1 has four main branches. These branches are the lacrimal, frontal, nasociliary, and meningeal branches. Each of these branches has multiple terminal branches, nearly all of which are general sensory. Some of these terminal branches are joined by visceral motor (parasympathetic) fibers from the ciliary ganglion and sympathetic fibers that travel along the long and short ciliary nerves.

The sensory component of V_2 has four main branches. These branches are the zygomatic, infraorbital, pterygopalatine, and meningeal branches. These branches carry mostly sensory information with few exceptions.

For example, the zygomatic nerve carries, for a short distance, postganglionic parasympathetic fibers from the seventh nerve that are on their way to the lacrimal gland. Other smaller branches, such as the greater and lesser palatine branches of the maxillary nerve, traverse the pterygopalatine ganglion without synapsing.

The sensory component of V_3 has five main branches. These branches are the buccal, auriculotemporal, lingual, inferior alveolar, and meningeal branches. Of note here is the fact that the chorda tympani fibers [special sensory (taste and visceral motor) parasympathetic] travel along the lingual nerve and at the level of the lingual nerve travels between the medial pterygoid muscle and the mandible.

The motor branch forms a common trunk with V_3, with which it travels for a short distance. The motor branch then divides into six motor branches. These branches are the medial pterygoid, masseteric, deep temporal, lateral pterygoid branches, and the nerves to the mylohyoid and anterior belly of the digastric. The nerves to the tensor tympani and tensor veli palatini are smaller branches of the medial pterygoid branch. It is of note here that the masticator (motor) nucleus receives input from the sensory branches of CN V and other sensory cranial nerves. For example, sensory input from the cornea (e.g., gently blowing air onto the cornea) activates the part of the motor nucleus that innervates the tensor tympani and results in contraction of this muscle.

Clinical Considerations for Otolaryngologists

- Facial fractures can result in facial anesthesia of the distal branches of the trigeminal nerve.
- The infraorbital nerve can be injured during the elevation and retraction of the cheek during a Caldwell-Luc approach to the maxillary sinus.
- Cavernous sinus lesions and/or thrombosis can result in anesthesia of V_1 and/or V_2. (V_2 does not always travel through the cavernous sinus.)
- Tic douloureux is a fairly common disorder that usually is due to a microvascular compression of the sensory root of the trigeminal nerve at the root entry zone within the posterior fossa. If the condition is not controlled with medical therapy, a microvascular decompression of CN V can be achieved via a retrosigmoid or retrolabyrinthine approach to the posterior fossa.
- Anesthesia of the face can be the result of a large acoustic neuroma. This is usually due to superior extension of the tumor and direct compression of the nerve. Additionally, CN V symptoms can be due to brainstem and resultant spinal tract compression by the acoustic neuroma.
- The blink reflex test has been used in the testing of patients with Bell's palsy and other causes of facial paralysis. CN V (supratrochlear or corneal stimulation) mediates the afferent portion of this reflex. The efferent portion is mediated via CN VII, resulting in blinking of the eye (orbicularis oculi muscle) (**Figs. 51–7** and **51–8**).

THE ABDUCENS NERVE (CN VI)

The abducens (CN VI) is a somatic motor (general somatic efferent) nerve that supplies the lateral rectus muscle of the eye. Contraction of this muscle results in an adduction (lateral horizontal) movement of the eye. Just as with the oculomotor and trochlear nerves, the function of the abducens nerve contributes to the coordintaed movement of each eye alone and to the combined synchronous movement of both eyes together.

The abducens nucleus is located in the pontine tegmentum close to the facial nerve nucleus. The motor fibers of the facial nerve loop around the abducens nucleus. Fibers leaving the abducens nucleus course ventrally, where they leave the brainstem as CN VI. CN VI courses within the posterior fossa, where it enters a dural fold (Dorello's canal) in the area of the petrous apex. It then pierces the dura and travels for a short distance before it enters the cavernous sinus, where it runs inferior and medial to the ophthalmic division of the trigeminal nerve (V_1). The nerve enters the orbit through the superior orbital fissure, then reaches and innervates the lateral rectus muscle.

An isolated (lower motor neuron) paralysis of the abducens nerve results in an inability to abduct the ipsilateral eye past the midline, resulting in diplopia on lateral gaze.

Clinical Considerations for Otolaryngologists

- The abducens nerve can be paralyzed in cavernous sinus lesions or thrombosis.
- Aneurysms of the posteroinferior cerebellar artery (PICA) and basilar or carotid arteries, and

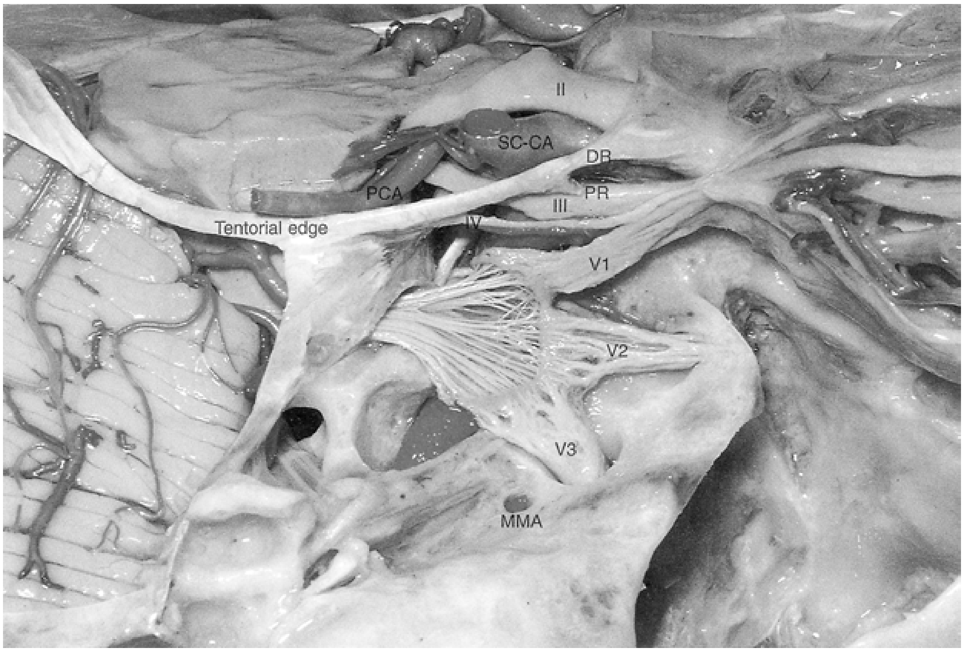


Figure 51–7 *The three branches of the trigeminal nerve (V_1, V_2, and V_3) are shown in relation to the middle meningeal artery and other neurovascular structures. DR, distal ring; MMA, middle meningeal artery; PCA, posterior cerebral artery; PR, proximal ring; SC-CA, supraclinoid carotid artery. (From Sen C, Chen CS, Post KD. Microsurgical Anatomy of the Skull Base and Approaches to the Cavernous Sinus. Stuttgart: Georg Thieme Verlag; 1997:49.)*

fourth ventricle lesions can cause abducens paralysis.

- Skull base fractures can result in abducens paralysis because of the proximity of the nerve to the petrous apex of the temporal bone.

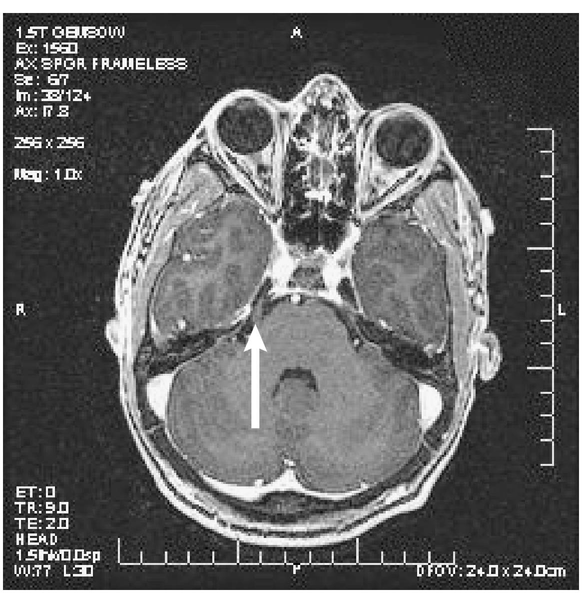


Figure 51–8 *Fifth nerve as it travels toward Meckel's cave, as seen on magnetic resonance imaging.*

- Petrous apicitis is characterized by infection of the petrous apex of the temporal bone, headache, and a CN VI palsy (Gradenigo's syndrome).
- Head trauma can result in facial paralysis and an associated abducens nerve palsy. In such cases, the lesion is typically central, involving the facial nerve motor fibers in the area where they loop around the abducens motor nucleus.

THE FACIAL NERVE (CN VII)

The facial nerve (CN VII) has four components:

- The first component is branchial motor (special visceral efferent). It supplies the muscles of facial expression, the stylohyoid, the posterior belly of the digastric muscles, and the buccinator, platysma, and occipitalis muscles.
- The second component is visceral motor (general visceral efferent). It provides innervation for the secretory functions of the lacrimal gland, submandibular and lingual glands, and mucous membranes of the nose and hard and soft palates.
- The third component is general sensory (general somatic afferent). It supplies the concha and a small area behind the ear. It also supplies a small

area of the posterior external auditory canal and possibly a small area of the lateral surface of the tympanic membrane.

- The fourth component is special sensory (special afferent). It carries taste sensation from the anterior two thirds of the tongue and hard and soft palates.

The signals for voluntary motion of the facial muscles originate in the motor cortex and are routed to the facial nuclei via the corticobulbar tracts. The nuclei that control the movement of the forehead receive bilateral innervation, whereas those controlling the movement of the lower face receive contralateral innervation. After synapsing in the motor nucleus of CN VII, the fibers leave the nucleus and loop around the nucleus of CN VI before leaving the brainstem to travel to and then through the internal auditory canal. The facial nerve travels ventral to the CN VIII bundle within the cerebellopontine angle. There it is joined by fibers from the nervus intermedius. These fibers are nonmotor fibers (see later). The facial nerve then enters the internal auditory canal, where it assumes a superior and anterior position (superior to the cochlear nerve and anterior to the superior vestibular nerve). The facial nerve leaves the internal auditory canal to enter the middle ear at the level of the geniculate ganglion. At that level, parasympathetics destined to reach the lacrimal gland branch off the nerve as the greater superficial petrosal nerve. The facial nerve then takes a posterior (horizontal) course to the level of its second bend (genu). At that level it gives off the branch to the stapedius muscle. It then curves inferiorly (vertical segment) and travels within the mastoid bone down to the stylomastoid foramen. Prior to reaching the stylomastoid foramen it gives off the chorda tympani branch. The main nerve then leaves the mastoid cavity at the stylomastoid foramen and immediately gives off branches to the stylohyoid and posterior belly of the digastric muscles. It also gives off a small branch that courses posteriorly to innervate the occipitalis muscle. The nerve then enters the substance of the parotid gland, where it divides into multiple branches that supply the muscles of facial expression. These branches are the temporal, zygomatic, buccal, mandibular, and cervical branches.

The visceral motor component of the facial nerve carries fibers. Its nucleus is the superior salivatory nucleus made up by scattered cells within the pontine tegmentum. The efferent fibers leaving the salivatory nucleus form the nervus intermedius that travels along the motor facial nerve within the cerebellopontine angle. The parasympathetic fibers of the nervus intermedius then join the main trunk of the facial nerve, but they later branch off at the levels of the greater superficial petrosal and the chorda tympani nerves. Those fibers that form the greater superficial nerve join up with the sympathetic fibers from the carotid plexus and together form the nerve of the pterygoid canal. The parasympathetic fibers of the nerve of the pterygoid canal synapse in the pterygopalatine ganglion before traveling to innervate the lacrimal gland, as well as glands within the nasal and sinus cavities. The parasympathetic fibers within the chorda tympani leave the facial nerve and the middle ear to join up with the mandibular portion of CN V (V_3). These fibers then travel with the lingual nerve and synapse within the submandibular ganglion before innervating the submandibular gland and the lingual and minor salivary glands.

The special sensory fibers carry taste sensation from the anterior two thirds of the tongue and the hard and soft palates. The cell bodies of these fibers are in the geniculate ganglion. (The parasympathetic fibers travel through the geniculate ganglion without synapsing.) The central axons travel from the geniculate ganglion to the brainstem, where they synapse within the nucleus solitarius, also known as the gustatory nucleus.

The general sensory fibers of the facial nerve also have their cell bodies within the geniculate ganglion. Axons from these general sensory cell bodies travel centrally within the nervus intermedius and travel within the spinal tract of the trigeminal nerve. They then synapse in the spinal portion of the trigeminal nucleus. From there their impulses radiate centrally to the contralateral sensory cortex.

Clinical Considerations for Otolaryngologists

- The facial nerve is a complex nerve that is of particular interest to otolaryngologists otologists, neurotologists, and head and neck surgeons.
- A detailed knowledge of facial nerve anatomy is essential for the safe performance of surgery in the face, neck, parotid gland, middle ear, inner ear and labyrinth, mastoid, internal auditory canal, and cerbellopontine angle.
- The facial nerve can be injured in fractures of the temporal bone. Decompression of the facial nerve following traumatic and idiopathic facial paralysis is a persistently hot topic of discussion.
- Facial nerve regeneration following facial paralysis is commonly associated with facial synkinesis

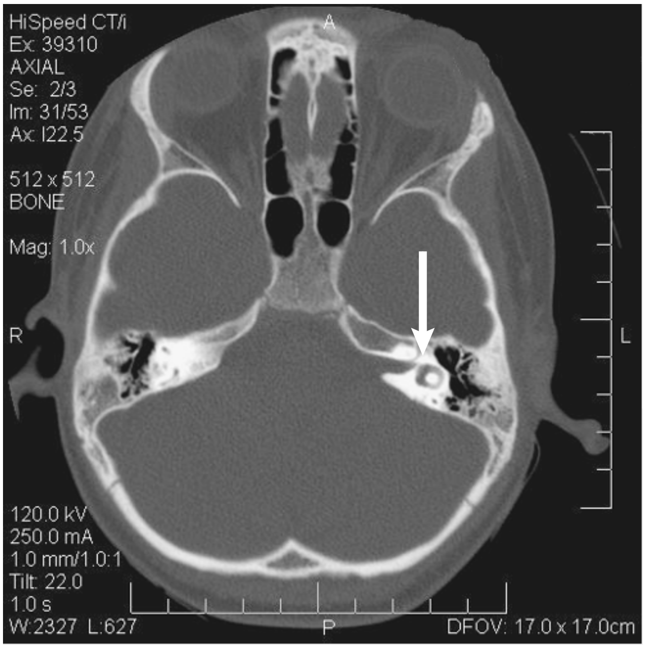


Figure 51–9 *Bill's bar (arrow) separating the facial nerve from the superior vestibular nerve (computed tomographic scan, axial cut).*

secondary to faulty rerouting of facial nerve fibers.

- The facial nerve can be rerouted for access to jugular bulb tumors and tumors in the petroclival area (**Figs. 51–9** and **51–10**).

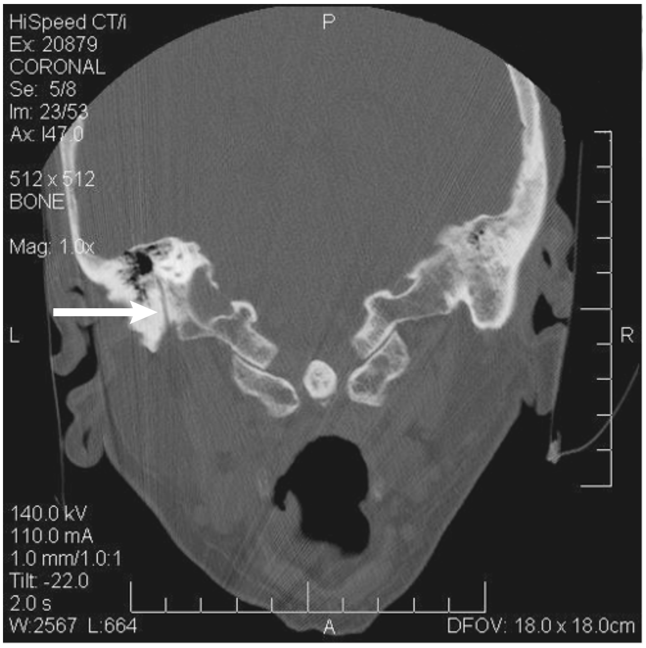


Figure 51–10 *Coronal computed tomographic scan view of the descending (mastoid) portion of the facial nerve canal within the mastoid bone (white arrow).*

THE VESTIBULOCOCHLEAR NERVE (CN VIII)

The vestibulocochlear nerve (CN VIII) is a special sensory (special afferent) nerve. It carries auditory information from the cochlea and vestibular information from the semicircular canals. It is worthwhile noting that there is strong evidence that CN VIII also carries efferent fibers that transmit information from the brain to the cochlea. The efferent system seems to modulate hearing by acting on the outer hair cells of the cochlea. The central connections of the efferent system have not been studied in depth.

The auditory portion of CN VIII originates within the cochlea at the modiolus and travels in the inferior anterior quadrant of the internal auditory canal to the brainstem. As it exits the internal auditory canal, it joins with the vestibular portion to form the vestibulocochlear nerve. Within the cerebellopontine angle, the fibers within the vestibulocochlear bundle continue to be largely segregated and usually separated by a cleavage plane until the fibers enter their respective nuclei at the brainstem.

The auditory neurons receive signals from the auditory hair cells. The signals are then transmitted to the ganglion cells of the auditory nerve. The ganglion cells are located within the modiolus of the cochlea. The signals are then projected centrally to the cochlear nuclei at the pontomedullary junction. There is a tonotopic distribution of the auditory hair cells, auditory nerve fibers, and cochlear nuclei. The central auditory pathways are less well understood. Most secondary auditory neurons send most of their axons to the contralateral lateral lemniscus. A smaller group of axons travels up the ipsilateral lateral lemniscus. The axons then travel to the inferior colliculus, where they synapse. The axons travel from the inferior colliculus to the medial geniculate body of the thalamus, where they synapse. The axons of the medial geniculate body neurons travel to the auditory cortex (transverse temporal gyrus and other areas with cortical auditory representation).

The vestibular portion of CN VIII originates within the posterior portion of the inner ear. Distally, it has two branches: the superior and inferior vestibular nerves. The superior nerve carries signals from the superior and lateral semicircular canals and utricle, and the inferior nerve carries signals from the posterior semicircular portion of the utricle and the saccule. These nerves travel within the superior and inferior posterior quadrants of the internal auditory canal. They then join the cochlear nerve in its course toward the brainstem.

The vestibular neurons receive signals from vestibular hair cells located within the semicircular canal ampullae and the maculae of the utricle and saccule. The ganglion cells for the primary vestibular neurons are located within a thickening of the vestibular nerve at Scarpa's ganglion. Axons from Scarpa's ganglion travel to the vestibular nuclear complex within the floor of the fourth ventricle. The secondary vestibular neurons send fibers to the cerebellum and down the ipsilateral spinal cord to form the lateral vestibulospinal tract. They also send neurons along the medial longitudinal fasciculus (MLF). The ascending fibers of the MLF terminate within the nuclei of CN III, CN IV, and CN VI. These fibers help coordinate the movement of the eyes during changes in posture. Vestibular axons within the descending portion of the MLF form the medial vestibulospinal tract and modulate the function of the lower motor neurons within the cervical spinal cord.

Clinical Considerations for Otolaryngologists

- The layout of the various nerves within the internal auditory canal (IAC) has major surgical implications. The IAC is generally divided into four quadrants. The facial nerve lies within the superior anterior quadrant, the cochlear nerve within the anterior inferior quadrant, and the vestibular nerves within the posterior quadrants superiorly and inferiorly. The quadrants are not equal in size. The superior and inferior halves of the IAC are separated by a horizontal bony partition called the transverse crest or the crista falciformis. The superior quadrants are separated by a vertical bony partition called the vertical crest or Bill's bar (named after William F. House, MD). The inferior quadrants are not separated by a bony partition. In other words the cochlear nerve and inferior vestibular nerves are not separated by bone. Bill's bar separates the superior vestibular nerve from the facial nerve.
- Tumors within the IAC are usually vestibular schwannomas (acoustic neuromas). Neuromas of the facial and cochlear nerves have been described. Meningiomas, lipomas, and other rarer tumors can also grow within the IAC.
- The auditory brainstem response waveforms reflect the various synapse stations that auditory signals generate on their way from the cochlea to the auditory cortex.
- The segregation and the cleavage plane between the vestibular and cochlear fibers within the vestibulocochlear nerve enable surgeons to achieve an effective selective vestibular nerve section within the cerebellopontine angle without the need to dissect and open the IAC (**Fig. 51–11**).

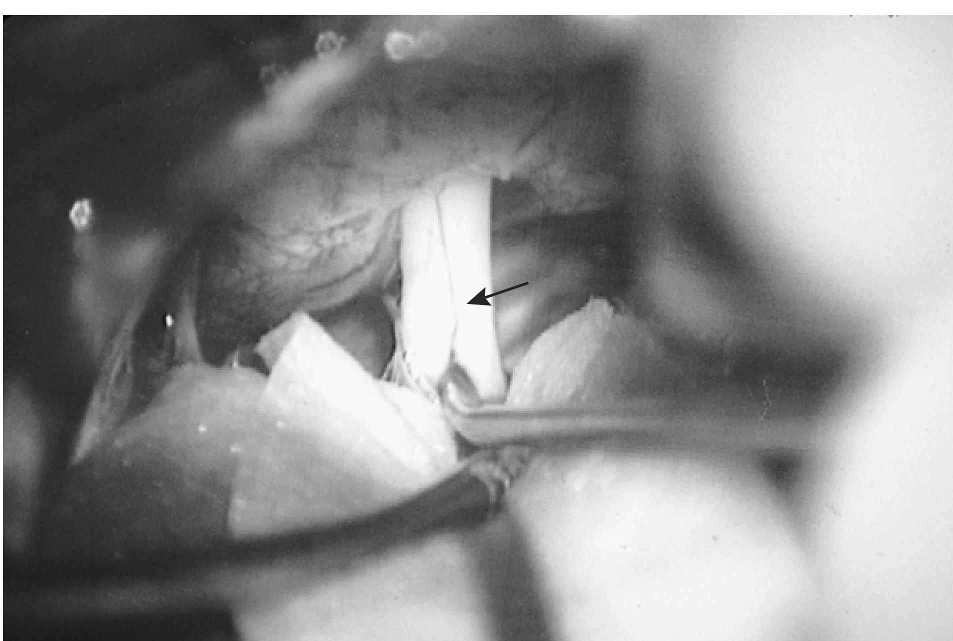


Figure 51–11 *Intraoperative view of the left eighth nerve in the posterior fossa. The arrow indicates the vestibular portion of the eighth nerve.*

THE GLOSSOPHARYNGEAL NERVE (CN IX)

The glossopharyngeal nerve (CN IX) has five components. The first component is branchial motor (special visceral efferent); it supplies the stylopharyngeus muscles. The second component is visceral motor (general visceral efferent) and supplies the otic ganglion, which sends fibers to stimulate the parotid gland. The third component is visceral sensory and carries sensations from the carotid body and carotid sinus. The fourth component is general sensory (general somatic afferent) and provides sensation from the posterior one third of the tongue, the external auditory canal, and the medial surface of the tympanic membrane. The fifth component is special sensory (special afferent) and carries taste sensation from the posterior one third of the tongue.

The glossopharyngeal nerve emerges as multiple roots from the medulla of the brainstem. These rootlets combine as they enter the jugular foramen as the glossopharyngeal nerve. The ganglia that mediate general, visceral, and special sensations lie within the pars nervosa of the jugular foramen.

The branchial motor axons travel down from the jugular foramen deep to the styloid process to reach the stylopharyngeus. The stylopharyngeus elevates the pharynx during swallowing and speech.

Special sensory axons carry taste sensation from the posterior one third of the tongue to the inferior ganglion within the jugular foramen. The taste fibers then travel with CN IX through the jugular foramen to enter the medulla and travel along the tractus solitarius to the nucleus solitarius. Taste sensation is eventually carried centrally to the inferior part of the postcentral gyrus, where the sensation of taste is perceived.

The parasympathetic fibers originate in the inferior salivatory nucleus in the medulla. The nucleus receives influencing innervation from the hypothalamus and the olfactory system. The parasympathetic fibers leave the medulla with the other components of CN IX and travel with these fibers through the jugular foramen.

The tympanic nerve is of special interest. It leaves the inferior ganglion and travels through the inferior tympanic canaliculus into the tympanic cavity, where it forms the tympanic plexus (Jacobson's nerve and plexus). The sensory fibers supply sensation to the middle ear and the eustachian tube, and the visceral motor fibers continue on and form the lesser petrosal nerve. This nerve then travels to the middle cranial fossa and lateral to the greater petrosal nerve. The lesser petrosal nerve fibers continue to descend through the foramen ovale. Immediately below the foramen ovale lies the otic ganglion, where the glossopharyngeal fibers synapse. Postganglionic fibers are then joined by fibers from the auriculotemporal nerve (branch of V_3) to supply the parotid gland.

Clinical Considerations for Otolaryngologists

- CN IX travels through the jugular foramen along with CN X and CN XI. Isolated CN IX lesions are rare. They are usually associated with lesions of CN X and CN XI (jugular foramen syndrome).
- CN IX is sensory to the soft palate. A unilateral decreased gag reflex is elicited when there is a lesion of the ipsilateral CN IX.
- Freye's syndrome or gustatory sweating occurs following parotidectomy surgery. Parasympathetic fibers that supply the transected parotid gland regrow and supply the sweat glands of the overlying skin. When salivation is stimulated, the sweat glands on the side of the parotidectomy are also stimulated. These parasympathetic fibers have their origins in the tympanic branch of CN IX. A tympanic neurectomy has been advocated for the control of gustatory sweating.
- Glossopharyngeal neuralgia is a rare condition characterized by pain in the throat radiating to the ear and neck.

THE VAGUS NERVE (CN X)

The vagus nerve (CN X) has four components. The first component is branchial motor (special visceral efferent) and supplies the striated muscles of the fourth branchial arch. These muscles are muscles of the pharynx (except the stylopharyngeus and tensor veli palatini muscles), the palatoglossus muscle, and the laryngeal muscles. The second component is visceral motor (general visceral efferent) to smooth muscles and glands of the pharynx and larynx. It also supplies the glands of the thoracic and abdominal viscera. The third component is visceral sensory (visceral afferent). It carries sensations from the larynx, trachea, esophagus, and thoracic and abdominal viscera. It also carries sensations from stretch receptors from the aortic arch and chemoreceptors in the aortic bodies. The fourth component is general sensory (general somatic afferent) and carries sensations from parts of the skin of the ear canal and lateral tympanic membrane and pharynx.

The vagus nerve is identifiable as several rootlets that emerge from the medulla of the brainstem. These rootlets combine as two main roots that enter the jugular foramen along with CN IX and CN X. Within the foramen CN X has two ganglia (superior or jugular and inferior or nodosum). The fibers leaving the inferior ganglion are joined by fibers from the nucleus ambiguus that had been traveling along CN XI. The vagus nerve travels within the carotid sheath and follows a different course on the right as compared with the left.

The branchial motor component is composed of fibers that originate in the cortex and travel with the corticobulbar tract down the nucleus ambiguus of the medulla. The lower motor neurons form about nine or 10 rootlets, some of which travel temporarily with CN XI. The other rootlets join into two main roots that enter the jugular foramen. The branchial motor fibers continue within the trunk of the vagus nerve. The first motor branch is the pharyngeal branch. It is the main motor branch to the pharynx. The next motor branch is the superior laryngeal branch. It supplies the inferior constrictor and the cricothyroid muscles. The third major branch is the recurrent laryngeal nerve. The right recurrent nerve loops around the subclavian artery in the neck and travels back up superiorly to supply the intrinsic muscles of the larynx except for the cricothyroid muscle. The left nerve loops around the aortic arch within the thoracic cavity. It then travels superiorly and innervates the laryngeal muscles on the left except the cricothyroid muscle.

The visceral motor branch has its parasympathetic nerve cell bodies within the dorsal nucleus of the vagus. The nucleus receives influencing stimuli from the hypothalamus, olfactory system, reticular formation, and the nucleus of the tractus solitarius. The vagal preganglionic axons activate ganglionic neurons that are secretomotor to the glands of the pharynx and larynx. They also send branches that join plexuses around the main blood vessels and organs within the thorax and abdomen.

The visceral sensory component carries the majority of sensations within the larynx through the internal branch of the superior laryngeal nerve. The recurrent laryngeal nerves carry sensation from the subglottic area. The axons of these neurons and those originating within the heart, stomach, and other viscera enter the medulla and descend and enter the nucleus of the tractus solitarius. From the nucleus, bilateral connections are made to the reticular formation, hypothalamus, and other cranial nerve nuclei at the brainstem. These connections enable the vagus to control multiple important reflexes relating to cardiovascular, respiratory, gastrointestinal function, and other important reflex responses.

The general sensory component carries sensation from the larynx, pharynx, skin of the ear, and tympanic membrane fossa. After synapsing within the jugular fossa, the fibers join other fibers carrying sensations from the meninges of the posterior cranial fossa and enter the medulla. The central processes then descend the spinal trigeminal tract and synapse in its nucleus. The proximal projections cross the midline and eventually reach the sensory cortex.

Clinical Considerations for Otolaryngologists

- The superior and recurrent laryngeal nerves are at risk during neck and thoracic surgery, such as carotid endarterectomy, thyroid surgery, anterior cervical disk surgery, and surgery on the aortic arch.
- Stimulation of the ear canal can result in coughing or more severe reflex responses because of stimulation of the sensory branch of the vagus nerve within the external auditory canal (Arnold's nerve).
- A high vagal injury can result in nasal regurgitation because of paralysis of the levator veli palatini muscle (**Fig. 51–12**).

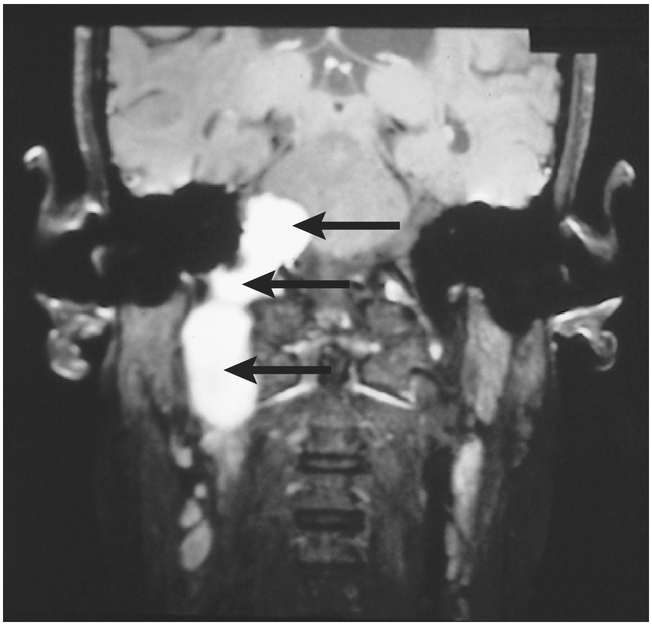


Figure 51–12 *Coronal magnetic resonance imaging view of a vagal nerve schwannoma with the intracranial (top arrow), jugular foramen (middle arrow), and neck (lower arrow) components of the tumor.*

Figure 51–13 *Surgical specimen of the jugular and neck portion of the tumor shown in **Fig. 51–12**. Note that the intracranial component had been resected separately.*

THE SPINAL ACCESSORY NERVE (CN XI)

The spinal accessory nerve (CN XI) is a motor nerve that supplies the sternocleidomastoid and the trapezius muscles. Some texts consider it to be branchial motor (special branchial efferent) because its nucleus is located in the ventral horn that is in line with the nucleus ambiguus. Other texts consider this nerve to be somatic motor (general somatic efferent) to the sternocleidomastoid and trapezius muscles. Still others consider it to be mixed somatic and branchial motor.

In addition, some texts consider the spinal accessory nerve as a purely spinal nerve and others as a mixed cranial and spinal nerve. This confusion arises from the fact that the lower motor neurons that travel within the spinal accessory nerve arise mostly from the (spinal) accessory nucleus, but they are joined by other fibers originating within the nucleus ambiguus. These latter fibers leave the spinal accessory nerve in the jugular foramen and join the vagus nerve; they could be considered to be the vagus nerve rather than spinal accessory fibers.

Be that as it may, the spinal accessory nucleus is located in lateral part of the anterior gray column of the upper five or six segments of the spinal cord. It receives input from cortical neurons via the descending corticospinal tract. Postsynaptic axons form the accessory nucleus rootlets that ascend into the posterior fossa via the foramen magnum. These axons are joined by other fibers originating from the nucleus ambiguus, as previously stated.

The spinal accessory rootlets enter the jugular foramen, which they exit into the neck as a single bundle that travels medially and posteriorly. The spinal accessory nerve then splits into two branches, one for each of the sternocleidomastoid and trapezius muscles. Here in the neck there is controversy as to whether fibers from the third and fourth cervical roots contribute to the motor function of the spinal accessory nerve.

Clinical Considerations for Otolaryngologists

- The spinal accessory nerve is at risk in neck surgery. It is routinely sacrificed in the classical form of radical neck dissection. Sectioning of the nerve results in a shoulder drop, head tilt, and winged scapula, with resultant neck pain and neck and shoulder stiffness (**Fig. 51–13**).

THE HYPOGLOSSAL NERVE (CN XII)

The hypoglossal nerve (CN XII) is a somatic motor (general somatic efferent) nerve that supplies all the intrinsic muscles of the tongue. It also supplies all the extrinsic muscles of the tongue except the palatoglossus muscle, which is supplied by the vagus nerve.

Axons from the motor cortex send corticobulbar fibers to the contralateral hypoglossal nucleus, where they synapse. The hypoglossal nucleus also receives influencing input from the nucleus of the tractus solitarius

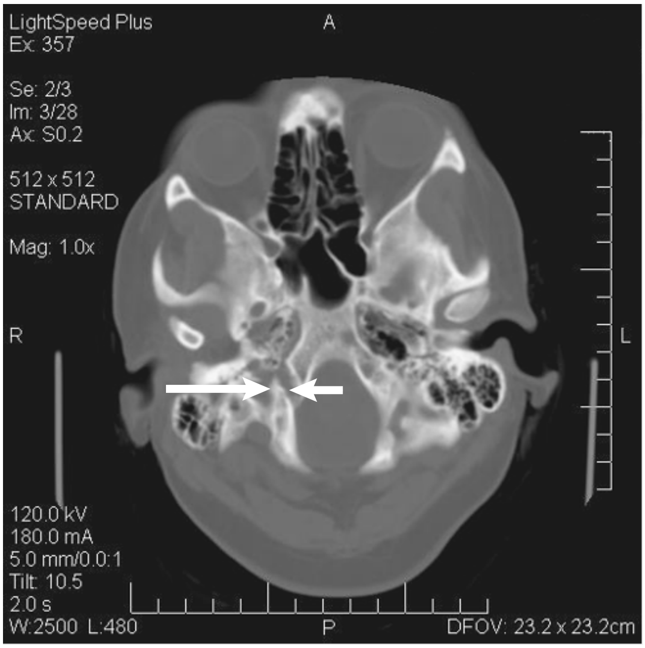


Figure 51–14 *Axial computed tomographic scan of the hypoglossal canal (short arrow) in close proximity to the jugular foramen (long arrow).*

in response to gustatory and tactile stimulation. The hypoglossal nucleus is located in the tegmentum of the medulla at the level of the dorsal nucleus of the vagus. Postsynaptic fibers leave the hypoglossal nucleus and travel as rootlets to the hypoglossal foramen. Within the neck the hypoglossal muscle travels medial to CN IX, CN X, and CN XI. It then moves laterally and curves anteriorly, medial to the digastric muscle, to supply the tongue muscles. During its course within the neck, the hypoglossal nerve carries fibers from cervical spinal nerve C_1. These fibers leave the hypoglossal nerve as the ansa hypoglossi. The ansa gets additional components from cervical spinal nerves C_2 and C_3.

The genioglossus muscles on both sides of the tongue help protrude the tongue in the midline. If the nerve to one side is paralyzed (lower motor neuron lesion), the tongue protrudes to the side of the paralysis.

Clinical Considerations for Otolaryngologists

- CN XII does not go through the jugular foramen. Tongue paralysis is not a component of the jugular foramen syndrome. Yet tongue paralysis is not uncommon in patients with glomus jugulare tumors. The tongue can be paralyzed by glomus tumors when these invade the nerve secondarily within the hypoglossal foramen and in the upper neck.
- Tongue paralysis can be due to isolated hypoglossal nerve schwannomas.
- Asymmetric protrusion of the tongue can be due to tongue carcinomas that limit the motion of the tongue on the involved side. Asymmetric protrusion should not be confused with hypoglossal paralysis (**Fig. 51–14**).

Suggested Readings

Janecka IP, Tiedemann K, eds. Skull Base Surgery Anatomy: Biology and Technology. Philadelphia: Lippincott-Raven; 1996

Lang J. Clinical Anatomy of the Posterior Fossa and Its Foramina. 1991

Sen C, Chen CS, Post KD. Microsurgical Anatomy of the Skull Base and Approaches to the Cavernous Sinus. Stuttgart: Georg Thieme Verlag; 1997

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 717.

1. Which cranial nerve does not travel through the jugular foramen yet is commonly paralyzed in patients with glomus jugulare tumors?
   A. Glossopharyngeal nerve
   B. Vagus nerve
   C. Spinal accessory nerve
   D. Hypoglossal nerve
2. Which cranial nerve is commonly at risk when the tentorium is severed or divided anteriorly?
   A. Oculomotor nerve
   B. Trochlear nerve
   C. Abducens nerve
   D. Optic nerve
3. Most acoustic neuromas arise from which of the following nerves?
   A. Superior vestibular nerve
   B. Acoustic portion of the eighth nerve
   C. Inferior vestibular nerve
   D. Nervus intermedius

Part V

THE ORAL CAVITY, TASTE, AND THE GLANDS OF THE NECK

Chapter 52

BASIC SCIENCE OF THE ORAL CAVITY AND GUSTATION

CHARLES P. KIMMELMAN

This chapter covers the anatomy, physiology, and pathology of the oral cavity and the gustatory system. It also includes a review of common testing for gustatory disturbances and treatment suggestions for gustatory disorders.

THE ORAL CAVITY

The oral cavity consists of the lips, buccal mucosa, tongue, maxillary and mandibular alveolar ridges and teeth, floor of the mouth, and soft and hard palates. Its primary function is to serve as the entrance of the alimentary tract, and as such to initiate the digestive process. It also functions as a secondary respiratory conduit, a site of sound modification for the production of speech, and a chemosensory organ. Because mammals are totally dependent on the mouth for their existence, the oral cavity is well endowed with sensory and reflex activities to facilitate ingestion and digestion and to promote the expulsion of noxious materials. Even minor disruptions in oral cavity function can place an individual in jeopardy of malnutrition and seriously impair communication.

FUNCTIONAL ANATOMY AND PHYSIOLOGY

Lips

The lips are musculomembranous portals of entry into the oral cavity that retain the food and liquid during mastication and deglutition. Within their submucosa are numerous minor serous and mucinous salivary glands that lubricate the mucosa and initiate digestion. The range of mobility of lip tissue is also critical to other functions, such as speech production, whistling and singing, the playing of wind and brass musical instruments, osculation, and expectoration. Human behavioral communication also depends on normal lip mobility (smile, smirk, pout, baring of teeth, etc.). The vermilion border is the transition from skin to lip mucosa. It is devoid of sweat and sebaceous glands and requires constant moistening. The numerous sensory nerve endings in the lips monitor the chemical composition, consistency, hydration, and temperature of incoming material to aid in mastication and deter entry of harmful substances.

Teeth

The teeth tear and grind ingesta, so that the latter can become saturated with liquid and digestive enzymes. The covering of the tooth is very hard enamel, the hardest substance in the body by virtue of its high composition (96–98%) of hydroxyapatite. This acellular material is produced by the ameloblast. Beneath the enamel is a mineralized (70% hydroxyapatite) organic matrix (collagen) called dentin. Odontoblasts produce dentin throughout life. The internal core, or pulp, contains loose connective tissue, nerves (branches of the alveolar nerves), and blood vessels to sustain the metabolic activity of the tooth. The cementum covers the unexposed root of the tooth and is a bonelike, avascular sheath. The periodontal ligament anchors the root to the alveolar process and creates a strong articulation with the socket known as a gomphosis. There are 32 adult (2 incisors, 1 canine, 2 premolars, and 3 molars per side per jaw) and 20 primary, or milk, deciduous teeth (2 incisors, 1 canine, and 2 deciduous molars per side per jaw). Incisors have one root and are blade shaped. Canines are one rooted and conical shaped, with a pointed tip. Premolars (bicuspids) have one or two roots with flattened crowns. The molars have three or more roots and have flat surfaces with prominent cusps for grinding and crushing.

Tongue

This remarkable organ consists of multiple muscles covered by a specialized mucous membrane. The muscle groups are intrinsic (wholly within the tongue: the inferior longitudinal, the superior longitudinal, and the intercalated transverse longitudinal) and extrinsic (genioglossus, hyoglossus, styloglossus, palatoglossus) and enable complex, yet precise, movements in all dimensions. Except for the palatoglossus, which is innervated by the cranial portion of the eleventh cranial nerve (CN XI), motor activity of the tongue is mediated by the hypoglossal nerve. The anterior two thirds of the tongue is covered by a nonkeratinized stratified squamous epithelium overlying a lamina propria that is organized into projections known as papillae. Most papillae are filiform, having tapered tips. Scattered among these are the much fewer mushroom-shaped fungiform papillae. The tongue's lateral edges are grooved by the booklike foliate papillae. The junction of the oral and pharyngeal (posterior one third) tongue is marked by an inverted V-shaped alignment of circumvallate papillae. The fungiform, foliate, and vallate papillae all support taste buds. Mounds of lymphoepithelial tissue known collectively as the lingual tonsil cover the base of the tongue.

The dexterity of the tongue allows it to perform tasks as diverse as an operatic aria and as mundane as the propulsion of a bolus into the pharynx. To perform these deft acts, an extensive sensory innervation is provided by the lingual nerves (branch of mandibular division of the trigeminal) to the anterior one third of the tongue and by the glossopharyngeal nerves to the tongue base.

Palate

The horseshoe-shaped, domed roof of the oral cavity has a firm, bony anchorage anteriorly (the hard palate) and a flexible, velum posteriorly (the soft palate). The hard palate has transverse ridges (rugae) to retain the food bolus. The soft palate is quite mobile and, in conjunction with the pharyngeal musculature, seals off the nasopharynx to prevent nasal regurgitation during deglutition and nasal air emission during speech. Within the lamina propria of both hard and soft palates are numerous minor salivary glands. The framework of the soft palate consists of a firm but thin aponeurosis formed by the tendons of the opposing tensor veli palatini muscles. The soft palate is elevated by the levator veli palatini, and the uvular muscle draws the central portion upward. The faucial arches comprise the palatoglossus and palatopharyngeus muscles, which tighten the oropharyngeal inlet, allowing it to grasp the food bolus during deglutition. Within the fauces are the palatine tonsils, which function as sentinels of the immune system. The deep tonsillar crypts permit antigen presentation to lymphocytes so that foreign matter can be recognized and inactivated by the adaptive immune system.

Salivary Gland

The salivary glands are exocrine glands derived from the ectoderm of the oral cavity. There are large, major glands (paired parotid, submandibular, and sublingual) and numerous minor glands throughout the lamina propria of the oral cavity (labial, buccal, palatine, and lingual). The parotid gland is shaped as a three-sided wedge situated between the temporal bone and mandibular ramus. Its horizontal excretory duct (Stensen's duct) loops over the lateral surface of the masseter muscle and enters the oral cavity opposite the maxillary second molar. The submandibular gland lies in the submandibular triangle and has an anterior extension above the free edge of the mylohyoid muscle. Its excretory duct (Wharton's duct) runs in the floor of the mouth along the medial border of the sublingual gland to pierce the surface of the floor of the mouth at the paramedian sublingual caruncle. The sublingual gland lies above the mylohyoid muscle and has multiple small ducts that open directly into the floor of the mouth. Salivary glands are compound tubuloalveolar structures of either serous or mucinous type. The parotid is entirely serous type, the submandibular is nearly all mucinous, and the sublingual contains both mucinous and serous gland types. Saliva contains the following electrolytes in decreasing concentration: sodium, potassium, chloride, bicarbonate, phosphate, calcium, and magnesium. Also present are various macromolecules, including blood group substances, mucins (lubricating glycoproteins), antivirals (cystatin), buffers (carbonic anhydrase, histatin), antibacterials [secretory immunoglobulin A (IgA), lactoferrin, lysozyme, peroxidases], neuroendocrine polypeptides (nerve growth factor, epidermal growth factor, renin, and kallikrein), and digestive enzymes (amylase).

The salivatory nuclei of the pons contain the first-order neurons controlling salivation. Some of the axons from these neurons travel with CN VII and leave the latter with the chorda tympani. The chorda enters the infratemporal fossa, joining the lingual nerve to synapse in the submandibular ganglion, thereafter sending postganglionic fibers to innervate the submandibular and sublingual glands. Other axons within CN VII leave with the greater superficial petrosal nerve, enter the vidian nerve, and synapse in the pterygopalatine ganglion. The postganglionic fibers enter the palatine nerves and activate minor glands in the palate. Preganglionic fibers also run with the glossopharyngeal nerve, leave it at the jugular foramen to ascend within the tympanic canaliculus, and traverse the middle ear as the tympanic plexus. The fibers continue intracranially as the lesser superficial petrosal nerve, then once more leave the cranial cavity through the foramen ovale and almost immediately synapse in the otic ganglion, whose postganglionic fibers enter the auriculotemporal nerve en route to innervate the parotid gland.

Salivary gland rates of secretion are $\sim$0.04 mL per major gland per hour. This rate diminishes significantly during sleep and is further affected by drugs, depression and anxiety, sympathetic overactivity, and level of hydration. Salivary gland activity requires cholinergic parasympathetic stimulation (neurotransmitters acetylcholine and vasoactive intestinal polypeptide) of muscarinic receptors at the acinar level. Sympathetic activity is weakly secretory and increases the proportion of macromolecules in saliva.

Intracellular secretory kinetics are mediated through second messengers. Cholinergic receptors are linked to G proteins that activate phospholipase C, resulting in the generation of inositol triphosphate from phosphatidyl inositol biphosphate. The latter releases calcium from the endoplasmic reticulum, leading to an opening of chloride channels in the luminal end of the cell and potassium channels at the basal end. As sodium ions subsequently enter the cell, secretion begins. Salivary proteins are released by the activity of sympathetic fibers, which stimulate adenyl cyclase to form cyclic adenosine monophosphate (cAMP) which activates protein kinase C, which drives secretory vesicles containing macromolecules to move toward and fuse with the cell membrane.

Primary secretion takes place in the acinar cells, and ductal cells modify the saliva by means of water and electrolyte resorption to increase potassium and decrease sodium concentration. Striated ductal cells synthesize the neuroendocrine polypeptides found in saliva.

Mastication

The act of chewing requires strong muscular forces to disrupt solid foods physically. The four muscles of mastication (internal and external pterygoids, temporalis, masseter) and the mylohyoid, geniohyoid, and genioglossus, along with the pull of gravity, provide a dynamic range of mandibular motion in all three dimensions. The articulation of the mandible with the skull, the temporomandibular joint, is thus exposed to significant compressive, torsional, and tractional forces. The joint is unique in its possession of avascular fibrous tissue over the articular surfaces, with an interposed meniscus partitioning the joint. The joint receives sensory innervation from the auriculotemporal branch of the mandibular division of the trigeminal nerve; hence the referral of pain to other branches of this nerve in the so-called temporomandibular joint syndrome (Costen's syndrome). This syndrome is caused by the action of excessive or unbalanced forces on the joint (e.g., dental malocclusion), joint trauma (stretching, external blow), or psychological perturbations (bruxism, stress-induced craniocervical muscle spasm).

Pathology of the Oral Cavity

Otolaryngologists are often consulted for oral cavity symptoms such as pain, oral dryness, presence of a mass, and bleeding. A working knowledge of the more common or serious pathological processes is critical to evaluate and manage patients with these complaints.

Oral Cavity Pain

This can be caused by external trauma (external force or entry of foreign body), internal trauma (abrasion of mucosa by teeth or food), infection of the gums and dental structures, ulceration (aphthous ulcers, autoimmune diseases, immunodeficiency) and neuralgia (Sluder's sphenopalatine neuralgia, trigeminal neuralgia, glossopharyngeal neuralgia).

Oral cavity ulcerations are a diverse group. Many of the lesions can be similar in appearance. Aphthae are oval, superficial, small (1–4 mm) lesions located over soft areas (buccal vestibule, undersurface of the tongue). They are more common in childhood and adolescence and usually last 4 to 7 days, healing without scarring. Giant lesions can occur in acquired immunodeficiency syndrome (AIDS). Autoimmune diseases form a rare, sometimes life-threatening, group of oral ulcerative lesions (pemphigus, lichen planus, pemphigoid, extraintestinal Crohn's disease, Wegener's granulomatosis). Biopsy of the lesion with immunohistochemical analysis by a pathologist well versed in oral pathology is mandatory.

Xerostomia

Xerostomia or dry mouth, is extremely common and is related to age, anticholinergic side effects of many drugs (e.g., antihistamines, antidepressants, antipsychotics, anxiolytics, antiparkinsons, antihypertensives, and diuretics), dehydration, radiation therapy, and many systemic disorders (e.g., hypothyroidism, estrogen deficiency, and amyloidosis). The evaluation of patients includes consideration of these possibilities with further evaluation as needed (autoantibody titers, minor salivary gland biopsy, hormone blood levels, etc.).

Sjögren's syndrome (autoimmune sialadenitis) deserves special mention because of its frequency and severity. The salivary glands become infiltrated with CD4 helper lymphocytes with a smaller contingent of plasma cells and macrophages. The normal gland architecture is gradually lost, and secretory elements degenerate into epimyoepithelial islands, a picture that is histologically described as the benign lymphoepithelial lesion. The abnormal glands enlarge and secondary lymphomas may appear after many years. Autoantibodies to salivary tissue can be detected in the blood and affirm the histological diagnosis.

Xerostomia allows the development of several important secondary oral disorders. Dental decay (caries) commonly develops with continued reduction in salivary flow because of the loss of antibacterial activity of saliva. The loss of defensive salivary components permits proliferation of cariogenic organisms. Xerostomia also permits proliferation of *Candida albicans*, leading to a severe stomatitis with white plaques, red patches, fissuring of the dorsal tongue, dysgeusia, and oral burning.

Masses of the Oral Cavity Structures

These can be of inflammatory, neoplastic, traumatic, toxic, or congenital origin. Infections of the minor salivary glands with secondary obstruction of salivary outflow may cause a mass effect. Sarcoidosis may involve the minor salivary glands. Amyloidosis infiltrates the tongue, turning it into a firm mass. Necrotizing sialometaplasia is a proliferative process resembling a neoplasm and resolves spontaneously. Odontogenic cysts may be infectious (apical), congenital (radicular), or neoplastic (ameloblastoma). Gingival hyperplasia can be induced by the action of drugs (phenytoin) and hormonal effects (epulis of pregnancy). Trauma to the oral mucosa and musculature leads to scarring and mass formation. It is not unusual to see a pedunculated mass (commonly called a fibroma) due to repeated masticatory injury to the lateral tongue or buccal mucosa. Ranulas are salivary extravasations of the sublingual glands usually induced by trauma. The resulting pseudocyst may remain in the space above the mylohyoid (sublingual) or may extend into the submandibular triangle (plunging), creating a neck mass. Congenital masses usually arise from fusion defects of the first branchial arch structures. Lymphangiomas and hemangiomas usually present in children and, if large, may cause serious airway and feeding difficulties. Benign neoplasms arise in the minor salivary glands (pleomorphic adenoma), mucosa (papillomas due to human papillomavirus), submucosa (granular cell tumors with overlying pseudoepitheliomatous hyperplasia), and nerves (schwannomas).

The oral cavity is one of the most common sites of origin of epithelial malignancies in the upper aerodigestive tract. Because of the prolonged contact of carcinogens with the mucosa of the floor of the mouth and ventral tongue, squamous cell carcinoma develops in this region in heavy users of tobacco or ethanol. Minor salivary gland malignancies (adenoid cystic carcinoma, mucoepidermoid carcinoma, acinic cell carcinoma, and undifferentiated carcinoma) are seen much less frequently. Primary and secondary neoplasms can more rarely develop in the bone of the maxilla and mandible (osteoma, myxoma, ameloblastoma, and osteogenic sarcoma).

GUSTATION

The senses of taste and smell, collectively known as chemosensation, have an impact on many processes within the body, such as nutrition, avoidance of danger, reproduction, and psychological homeostasis. Because the end-organs mediating these modalities are within the field of otolaryngological interest, specialists in this area should be familiar with the evaluation and management of chemosensory disorders.

Gustatory disorders are rare when compared with olfactory disturbances. When patients complain of a "taste" problem, they usually have an alteration in flavor perception due to an olfactory impairment. Flavor is defined as the aggregate of multiple sensations during mastication: olfaction, gustation, audition, and tactile sensations (touch, temperature, and pain). Nevertheless, true gustatory dysfunctions are occasionally encountered.

Anatomy and Physiology

Gustatory receptor cells are packed like sections of a grapefruit within taste buds. The latter contain ~100 receptor cells, which have an apical portion in contact with the bud inlet, the taste pore. Microvillae-bearing receptor molecules project into the pore. Analogous with the olfactory system, there is a layer of basal cells to reconstitute lost sensory receptor cells.

Unlike olfactory receptor cells, the taste receptors are not neurons. They receive a bouton from a sensory dendrite, whose cell body is located in the sensory ganglia of the facial (for taste buds on the anterior two thirds of the tongue), glossopharyngeal (for taste buds on the tongue base and pharynx), and vagus nerves (for taste buds in the larynx and esophagus).

Taste buds are harbored on the tongue, soft palate, pharynx, and larynx. Most are found on the tongue, where they are contained within specialized papillae. On the anterior two thirds of the dorsum of the tongue, each fungiform (mushroom-shaped) papilla contains ~20 taste buds. Laterally, the foliate (leaf-shaped) papillae support hundreds of buds in their grooves. The 12 or so circumvallate papillae are arranged in an inverted V formation at the junction of the anterior two-thirds and posterior third of the tongue. Each circumvallate papilla bears hundreds of taste buds in its moat. Taste buds present in other regions are scattered over the mucosal surface.

Gustatory perception begins with the solubilization of tastant molecules in saliva. When the molecules enter the taste pore, they interact with receptor proteins within the cell membrane of the microvillae or directly with ion channel structures themselves. Sweet and bitter tastants bind to receptor molecules that, in a manner similar to olfaction, activate various G proteins (one of which has been dubbed "gustducin" in an analogy to the transducin G protein in the visual pathway). G proteins activate enzymes such as adenyl cyclase and phospholipase C, which generate the second messengers: cAMP and inositol triphosphate. Sweetness transduction involves cAMP, and bitter transduction uses both second messengers. Receptor cell activation follows the opening of potassium channels in the case of sweet stimulation and intracellular release of calcium in the case of bitter tastants. In the case of salty stimuli, sodium ions enter the cell through ion channels, eventually causing depolarization and the entry of calcium with release of neurotransmitter vesicles. Acid taste is activated by the entry of hydrogen ions, which depolarize through their direct entry, their blockade of potassium channels on microvillli, or by allowing other cations to enter the cell.

The axons of the sensory neurons, which innervate the receptor cells, synapse within the tractus solitarius of the brainstem. The second-order neurons reside in the nucleus of the tract and send their axons centrally, with side sprouts sent to various autonomical centers in the brainstem. The axons of the solitary tract neurons proceed rostrally within the central tegmental tract and synapse in the medial part of the ventrobasal thalamus, which in turn projects to the postcentral gyrus.

The four main taste qualities are sweet, sour, salty, and bitter. Some studies have shown differential sensitivity for these qualities on different portions of the tongue, but such representations are meant to display slight variations in threshold detection at the various sites. A fifth quality, dubbed umami after the Japanese term for the taste of glutamate, is also recognized. It has been shown that receptor cells are preferentially, but not exclusively, stimulated by one of the four basic taste qualities. This has led neuroscientists to the conclusion that the overall pattern of neural stimulation is responsible for taste recognition.

Pathology of the Gustatory System

Malfunction of the gustatory apparatus can be a result of many causes:

- Xerostomia: Insufficient saliva or excessive salivary dehydration will impair transport of tastants to the buds. Saliva also contains trophic substances, which sustain the integrity of the receptor cells.
- Radiation therapy: Ionizing radiation injures or destroys receptor cells.

- Drug effects: In addition to anticholinergic drying effects, many drugs alter receptor cell function so that inappropriate messages are sent to the brain. Often, the resulting dysgeusia has a metallic character.
- Surgery: Excessive removal of tongue mucosa will diminish the sense of taste. Extensive oral surgery, such as maxillary-mandibular advancement and implants, may affect afferent innervation of taste buds. Middle ear surgery puts the chorda tympani nerve at risk.
- Infection: As in olfaction, viruses are felt to play a role in some cases of gustatory dysfunction. Oral cavity infections from bacteria (gingivitis, periodontitis) and fungi (candidiasis) may also adversely affect taste. Oral prostheses, such as bridges and implants, can become infected as well.
- Autoimmune disorders such as pemphigus and Sjögren's syndrome can impair taste.
- Gastroesophageal reflux can distort the sense of taste.
- Neoplasia: Neoplasms of the oral cavity can destroy the taste buds.
- Nutritional disorders, such as vitamin A, B complex, and C deficiency, can adversely affect gustatory function. Glossitis may result from iron deficiency as well.
- Endocrine metabolic: Various hormonal disturbances, such as hypothyroidism and diabetes mellitus, and metabolic disorders, including uremia and hepatic disease, may perturb gustation.
- Electrochemical effects engendered by galvanism induced by differential activity gradients of the metallic elements within oral cavity fillings and prostheses can give rise to taste auras.

Gustatory Evaluation

A complete head and neck history and physical examination are mandatory to identify abnormalities and risk factors for gustatory disturbances, as already listed. The physical examination should focus on the oral cavity (condition of the tongue and its papillation, gingiva, dental structures, buccal and palatal mucosa, quantity and quality of saliva, and status of the ductal orifices of the major glands), the neck (salivary glands, masses, scars, injuries, and congenital anomalies), the nose and nasopharynx, and the larynx.

Laboratory analysis of microbial flora (looking for aerobic, anaerobic, or fungal pathogens) and biopsy of oral cavity lesions are important. Blood studies for autoimmune disease, anemia, inflammation, hormonal dysfunction, and biochemical derangements are helpful. Radiographic evaluation of the neck, skull base, and central nervous system should be dictated by the findings of the history and physical examination.

Psychophysical testing using threshold or suprathreshold measures for specific tastants representing the four taste qualities is commonly done. The stimuli can be presented in increasing concentrations spanning the expected threshold in log increments. If abnormalities are found, a single staircase threshold detection is performed. If a specific region of the tongue is abnormal, regional or quadrant testing can be performed. Electrogustometric techniques stimulate taste receptors directly and are useful for obtaining threshold values. Injection of tastants into the venous system can be employed to ascertain taste function.

Treatment of Gustatory Disorders

Reversal or correction of the underlying etiology should be pursued. Removal of anticholinergic and other offending drugs and topical agents is mandatory. Patients who smoke, must stop. Restoration of oral cavity health is paramount: mucosal, tongue, dental, and gingival disease should be identified and ameliorated. Adequate salivary flow is a sine qua non, and salivary enhancement with procholinergic and hydrating agents is usually required. Elimination of nutritional deficiencies and hormonal disturbances is critical. In all patients, thorough chewing gives adequate time for the release of sapid molecules and prolongs their interaction time with receptors. Rotation of foods throughout a meal reduces sensory fatigue, and varying textures can add to the enjoyment of eating. Flavor enhancers also may help.

Suggested Readings

Atkinson J, Baum B. Salivary enhancement: current status. J Dent Educ 2001;65:1096–1101

Burn-Murdock R. Formation of Saliva. Available at: www.umds.ac.uk/physiology/rbm.1C8.htm

Greenspan D. Xerostomia: diagnosis and management. Oncology 1996;10(Suppl):7–11

Kimmelman C. Disorders of Taste and Smell: A Self-Instructional Package. Alexandria, VA: American Academy of Otolaryngology Head and Neck Surgery Foundation; 1996

Smith D, Margolskee R. Making sense of taste. Sci Am 2001; 284(3):32–39

Smith D, St. John S. Neural coding of gustatory information. Curr Opin Neurobiol 1999;9:427–435

Self-Test Questions

For each question select all the correct answers from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. The function(s) of the oral cavity include(s)
   - A. Initiation of digestion
   - B. Modulation of voice
   - C. Vibrato vocalization
   - D. Airway protection
2. The tongue is
   - A. An organ of digestion
   - B. An organ of chemosensation
   - C. A muscular organ
   - D. A part of the immune system
3. The salivary glands are adversely affected by
   - A. Many pharmaceutical agents
   - B. Radiation therapy
   - C. Autoimmune disease
   - D. Lymphoma
4. Salivary secretion is dependent upon
   - A. "Second messenger" biochemical processes
   - B. Cholinergic stimulation
   - C. Olfactory stimuli
   - D. Circumvallate papillae
5. Taste buds are present in which structure(s)?
   - A. Tongue
   - B. Palate
   - C. Gingiva
   - D. Epiglottis
6. Which of the following is true?
   - A. Sour tastants require intramembranous protein receptors for gustatory transduction.
   - B. Cyclic adenosine monophosphate is a second messenger for bitter stimuli.
   - C. Taste bud receptor cells send their axons with the chorda tympani nerve.
   - D. Filiform papillae contain taste buds on their apical surface.
7. When a patient complains of a loss of taste, the clinician should
   - A. Tell the patient that taste is intact, but the flavor sense is lost
   - B. Test olfactory ability
   - C. Perform a taste bud biopsy
   - D. Search for possible xerostomia
8. Gustatory coding strategies involve
   - A. Neural integration
   - B. Collective input from taste receptor cells with varying sensitivity to tastants
   - C. Differential taste quality sensitivity over different parts of the tongue
   - D. Input from umami receptors

Chapter 53

MORPHOPHYSIOLOGY OF THE SALIVARY GLANDS

RICHARD J. WONG AND GREGORY W. RANDOLPH

ANATOMY

PAROTID GLAND

The parotid gland is the largest of the paired major salivary glands. It is a triangular-shaped gland located in the preauricular region, deep to the facial skin and subcutaneous tissues that compose its lateral boundary. The parotid is bounded superiorly by the zygomatic arch, posteriorly by the external auditory canal and mastoid, posteroinferiorly by the obliquely running digastric and sternocleidomastoid muscles, and anteriorly by the masseter muscle. The medial or deep boundary is composed of the ramus of the mandible anteriorly, and the styloid process and parapharyngeal space posteriorly (**Fig. 53–1**).

The parotid gland is shaped like an asymmetrical dumbbell. The larger, wider portion of the gland lies lateral to the constriction formed by the ramus of the mandible and digastric muscle. The smaller, deeper portion of the parotid gland lies medial to this constriction adjacent to the parapharyngeal space. The plane between the superficial and deep lobes is defined by the facial nerve as it exits the stylomastoid foramen and courses anteriorly through the gland, branching to innervate the muscles of facial expression.

The parotid duct, or Stensen's duct, arises from the anterior border of the gland 1.5 cm inferior to the zygoma and runs anteriorly across the masseter muscle and buccal fat pad. The duct then dives deep, coursing medially through the buccinator muscle and emptying intraorally adjacent to the second maxillary molar. The duct's course is approximated by a line drawn from the external auditory canal meatus to a point just superior to the oral commissure.

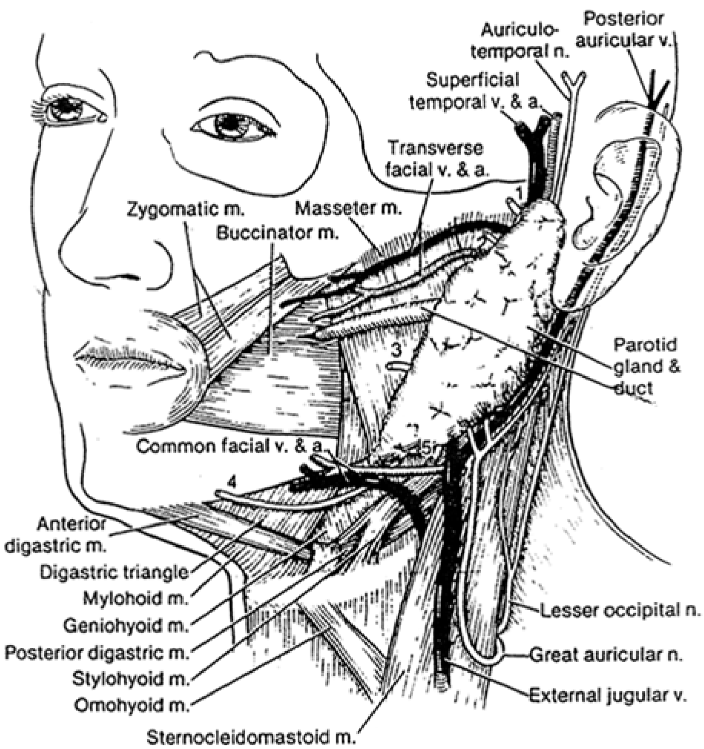


Figure 53–1 *The parotid gland and its surrounding structures. (From Graney DO, Jacobs JR, Kern RC. Salivary gland anatomy. In: Cummings CC, ed. Otololaryngology—Head and Neck Surgery. St. Louis: Mosby 1999:1201–1209. Reprinted with permission.)*

The fascia overlying the parotid gland is continuous with the superficial layer of the deep cervical fascia, which also envelops the adjacent sternocleidomastoid and masseter muscles. This robust fascial layer extends intraglandularly with multiple septae, creating a relatively inelastic capsule both around and through the parotid gland. This inelastic capsule prevents fluctuance from being present on physical exam in cases of parotid abscess.

Facial Nerve

The facial nerve emerges from the base of the skull at the stylomastoid foramen, between the styloid process and mastoid tip. It exits just anteroinferior to the tympanomastoid suture line and medial to the tragal "pointer," which is a projection of conchal cartilage. The depth of the nerve is approximated by the level of the digastric muscle, lateral to the styloid process. The main trunk initially gives off small branches to the posterior auricular muscle, the posterior belly of the digastric, and the stylohoid muscle. The facial nerve then courses anteriorly through the parotid gland and becomes progressively more superficial. It divides at a point termed the pes anserinus into its two main divisions: the upper temporofacial and lower cervicofacial divisions. These divisions then subdivide into five main branches. The temporofacial division gives rise to the temporal, zygomatic, and buccal branches, and the cervicofacial division leads to the buccal, marginal mandibular, and cervical branches. There may be multiple small connections between these branches as well as many variations in branching patterns. The nerve branches continue distally through the parotid gland to innervate the muscles of facial expression. A major goal in parotid surgery is often the identification and preservation of the facial nerve trunk and its branches.

Greater Auricular Nerve, Auriculotemporal Nerve, and Ninth Cranial Nerve

The greater auricular nerve is the largest branch of the cervical plexus. It emerges from the posterior border of the sternocleidomastoid muscle and reflects over its lateral surface to course in a superior direction. The nerve passes through the posterior parotid tail toward the pinna and ear lobule, where it supplies skin sensation.

The auriculotemporal nerve is a branch of the mandibular division of the trigeminal nerve. It courses anterior to the external auditory canal, and travels superiorly with the superficial temporal artery and vein to provide sensation to the scalp. Parasympathetic nerve fibers controlling secretory stimulation to the parotid travel with the auriculotemporal nerve. These secretomotor fibers originate in the inferior salivary nucleus in the pons and form part of the ninth cranial nerve (CN IX). Fibers branch off in the tympanic cavity as Jacobson's nerve and continue anteriorly as the lesser petrosal nerve before synapsing at the otic ganglion. The postganglionic nerve fibers then join with the auriculotemporal nerve

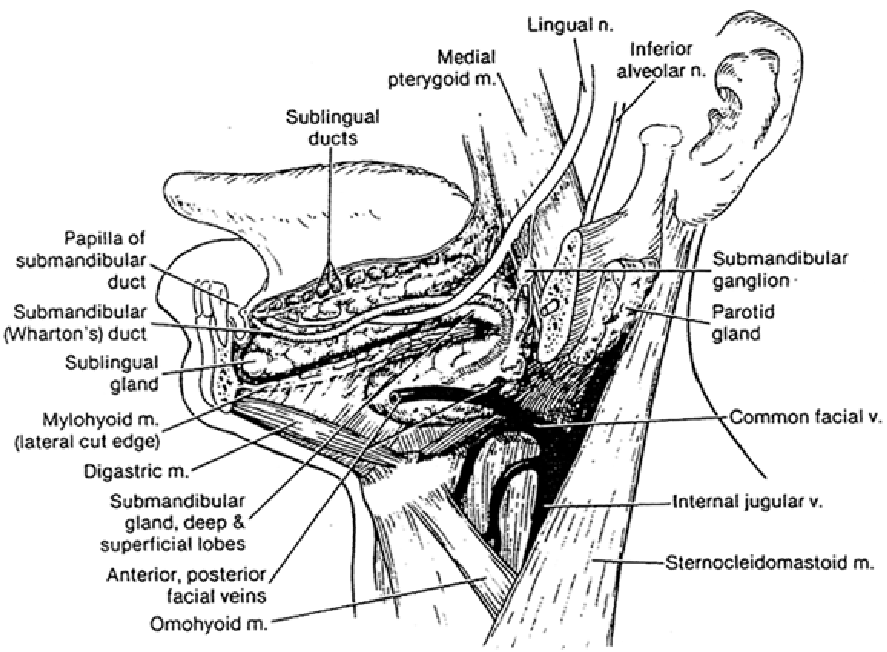


Figure 53–2 *The submandibular gland and its surrounding structures. (From Graney DO, Jacobs JR, Kern RC. Salivary gland anatomy. In: Cummings CC, ed. Otololaryngology—Head and Neck Surgery. St. Louis: Mosby 1999:1201–1209. Reprinted with permission.)*

and disperse into the body of the parotid gland. The postganglionic CN IX fibers are responsible for stimulation of parotid secretion, and the interruption of these nerve fibers may lead to gland atrophy. In some cases of parotidectomy these CN IX fibers are aberrantly reinnervated to the skin, causing gustatory sweating or Frey's syndrome.

Parotid Vascular Supply

The blood supply to the parotid gland comes from the external carotid artery. The external carotid artery branches into the internal maxillary artery and superficial temporal artery at the level of the mandibular condyle. The superficial temporal artery then gives rise to the transverse facial artery, which parallels Stensen's duct and supplies the parotid gland and upper quadrant of the face.

The superficial temporal vein joins the maxillary vein to form the retromandibular or posterior facial vein, which runs caudally within the parotid gland, deep to the facial nerve. The retromandibular vein emerges from the parotid gland to join the anterior facial vein. This union forms the common facial vein, which then empties into the internal jugular vein. The posterior facial vein also joins more superficially with the postauricular vein to empty into the external jugular vein.

Submandibular Gland

The submandibular gland is the second largest salivary gland. It lies in the submandibular triangle, a region bounded by the lower border of the mandible, the anterior belly of the digastric, and the posterior belly of the digastric muscle (**Fig. 53–2**). The floor of the submandibular triangle is composed of the mylohyoid muscle anteriorly and the hyoglossus muscle just deep to it. The submandibular gland is C-shaped and wraps around the posterior edge of the mylohyoid muscle, which defines the deep and superficial portions of the gland.

The submandibular duct, or Wharton's duct, emerges from the deep lobe of the gland and courses anteriorly between the mylohoid and hyoglossus muscles. It continues through the floor of the mouth to empty intraorally just lateral to the frenulum of the tongue. Both the hypoglossal nerve, which provides tongue motor innervation, and the lingual nerve, which provides tongue sensation, also course within the submandibular triangle deep to the mylohyoid muscle adjacent to Wharton's duct. All three structures may be exposed in the floor of the submandibular triangle by retraction of the mylohyoid muscle anteriorly and retraction of the submandibular gland posteriorly. The facial artery courses from the external carotid artery, through the deep portion of the submandibular gland, and over the inferior edge of the mandible. This artery must be ligated twice during excision of the submandibular gland.

Parasympathetic innervation of the submandibular gland is provided by the chorda tympani nerve, a branch of the facial nerve. The chorda tympani nerve branches off from the descending mastoid portion of the facial nerve and joins the course of the lingual nerve. In the submandibular triangle, the nerve fibers of the chorda tympani synapse at the submandibular ganglion, which lies between the lingual nerve and the submandibular gland. The postganglionic fibers then innervate the submandibular gland and stimulate secretory activity.

The superficial layer of the deep cervical fascia splits to envelop the submandibular gland. The marginal mandibular branch of the facial nerve courses within this fascia, superficial to the facial vein, and should be protected in cases where the submandibular gland is surgically resected. Ligation and elevation of the facial vein and submandibular fascia are techniques used to protect the marginal mandibular nerve during submandibular gland excision.

Sublingual Gland

The sublingual gland is the smallest of the major paired salivary glands. This almond-shaped gland lies in the floor of the mouth just deep to the mucosa, near the mandibular symphysis, and superior to the mylohyoid muscle. It opens intraorally through 8 to 20 small ducts from the superior surface of the gland called the ducts of Rivinus. The sublingual gland is not contained by a fascial capsule.

The parasympathetic nerve supply to the sublingual gland, like the submandibular gland, is from the chorda tympani nerve, which travels with the lingual nerve. The facial artery carries the sympathetic nerve supply from the cervical ganglion. The arterial supply to the sublingual gland comes from the sublingual branch of the lingual artery and the submental branch of the facial artery.

Minor Salivary Glands

There are ~600 to 1000 minor salivary glands. These are most concentrated in the buccal, palatal, labial, and lingual regions, although they also line the entire oral cavity. Each gland empties saliva into the oral cavity through its own single duct. Most minor salivary glands are innervated by the parasympathetics of the lingual nerve, although palatal glands are controlled by nerve fibers from the sphenopalatine ganglion. The blood supply to these minor salivary glands corresponds to the supply of the region in which they are located.

PHYSIOLOGY

Function of Saliva

Saliva has a wide variety of functions in the oral cavity. Saliva provides lubrication for speech and swallowing, mechanical cleaning, mucosal and dental protection, taste mediation, enzymatic digestion, and a medium for numerous proteins and immunoglobulins. The secretion of saliva is a complex process. The autonomic nervous system stimulates salivary fluid and protein secretion by the acinar cells. The saliva is then transported and modified by the ductal system prior to excretion into the oral cavity.

Autonomic Control of Salivation

The physiological control of salivation is maintained by the autonomic nervous system. Although stimulation of either the parasympathetic or the sympathetic system results in salivary secretion, the effects of the parasympathetic system are far more dominant. Interruption of the parasympathetic system has been shown to result in salivary gland atrophy, whereas interruption of the sympathetic system does not significantly affect gland function. The sympathetic nervous system is a modulator of the saliva composition and stimulates protein secretion, and the parasympathetic system is a stimulant for fluid secretion.

The autonomic nervous system releases neurotransmitters from adjacent nerve endings that bind to receptors on salivary gland acinar cells and stimulate salivary secretion. Salivary acinar cells are polarized and are linked by tight junctions that separate the apical, or luminal, side from the basolateral side. Sympathetic stimulation occurs through norepinephrine, and parasympathetic stimulation occurs through acetylcholine. These neurotransmitters are released and act on receptors located on the acinar cell's basolateral membrane. Norepinephrine binds to α- or β-adrenergic receptors, and acetylcholine binds to muscarinic cholinergic receptors. The cell's secretory products are released on the apical side.

There are two main types of neuroeffector arrangements between the autonomic nerve axons and salivary cells. In the epilemmal arrangement, the nerve axon lies outside the salivary cell basement membrane; in the hypolemmal arrangement, the axon penetrates the basement membrane. Neuroactivation occurs when sufficient neurotransmitter is released to activate the adjacent salivary cells. This activation is dependent on the distance between axon and cell and receptor sensitivity.

Secretory Unit

The basic secretory unit of all salivary glands consists of the acinus, the intercalated duct, the striated duct, and the interlobular duct (**Fig. 53–3**). The acinus is the site of all fluid generation and 85% of the exocrine protein secretion by the gland. The salivary fluid and proteins are secreted by the acinar cells into a central lumen. There are two types of secretory acinar cells. The serous type produces a watery secretion, and the mucinous type

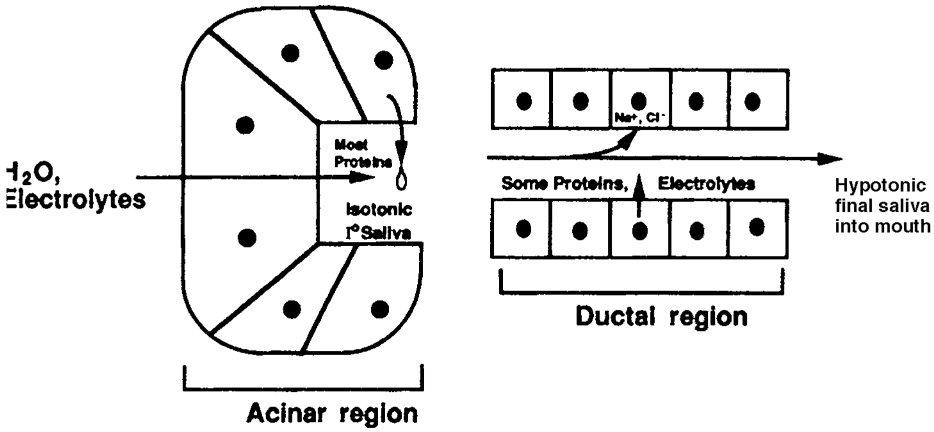


Figure 53–3 *Schematic depiction of the basic secretory unit. Saliva formation first occurs in the acinar region with the generation of an isotonic fluid rich in proteins. This fluid then passes through the ductal region, where sodium and chloride are resorbed, and some potassium, bicarbonate, and proteins are secreted. The final saliva entering the mouth is markedly hypotonic. (From Baum BJ. Principles of saliva secretion. Ann NY Acad Sci 1993; 694:17–23. Reprinted with permission.)*

produces a viscous secretion. The parotid gland contains serous acinar cells, which produce predominantly amylase, the sublingual gland contains acinar cells, which produce mucins, and the submandibular gland contains a mixture of both cell types. Minor salivary glands may contain acini of serous, mucinous, or mixed cell types.

Myoepithelial cells are interposed between the acinar cells and their peripheral basal membrane. The myoepithelial cells form a contractile network that expels saliva from the acini into the ductal system. Saliva formed in the acini passes into the intercalated ducts, a relatively short system lined by a single layer of cuboidal epithelium. The fluid then continues through the columnar-lined striated ducts, through the interlobular ducts, and empties into the oral cavity. The ductal system not only serves as a transport system but also modifies the electrolyte composition of saliva.

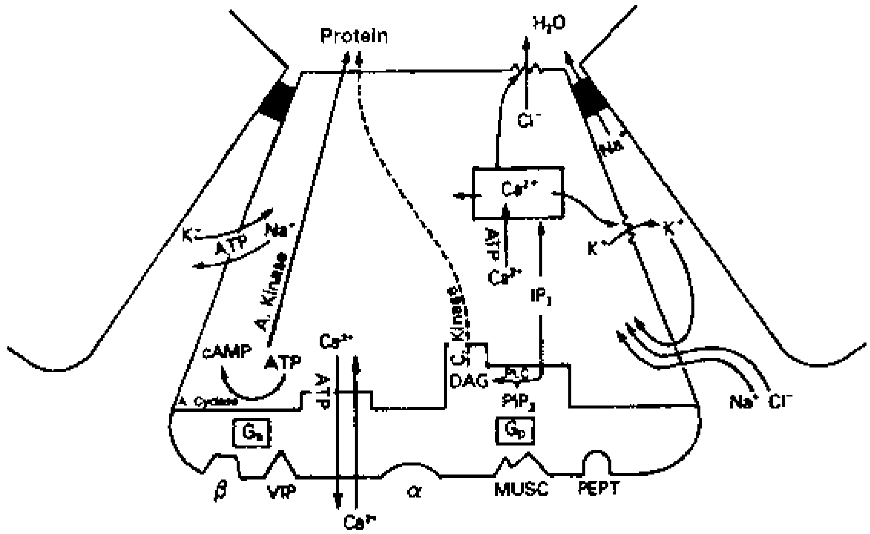
Mechanism of Acinar Cell Protein and Fluid Secretion

Stimulation of the acinar cell by norepinephrine released by the sympathetic nervous system leads to an increase in salivary protein secretion. The acinar cell membrane β-adrenergic receptors are triggered, and intracellular cyclic adenosine monophosphate(cAMP) rises (**Fig. 53–4A**). Next, A kinase is activated, which subsequently phosphorylates or dephosphorylates several cellular proteins. The exact details within this chain of events remain to be elucidated, but they ultimately lead to an increase in acinar protein secretion. Proteins are synthesized in the rough endoplasmic reticulum and are segregated into the cisternal space before transportation to the Golgi complex for processing. The proteins are then concentrated within secretion granules, which eventually discharge their contents through the apical membrane of the acinar cell through the process of exocytosis. In the parotid gland salivary amylase is predominantly secreted; in the submandibular gland mucin-glycoproteins are highly produced. Many other proteins are also present in saliva (see later discussion).

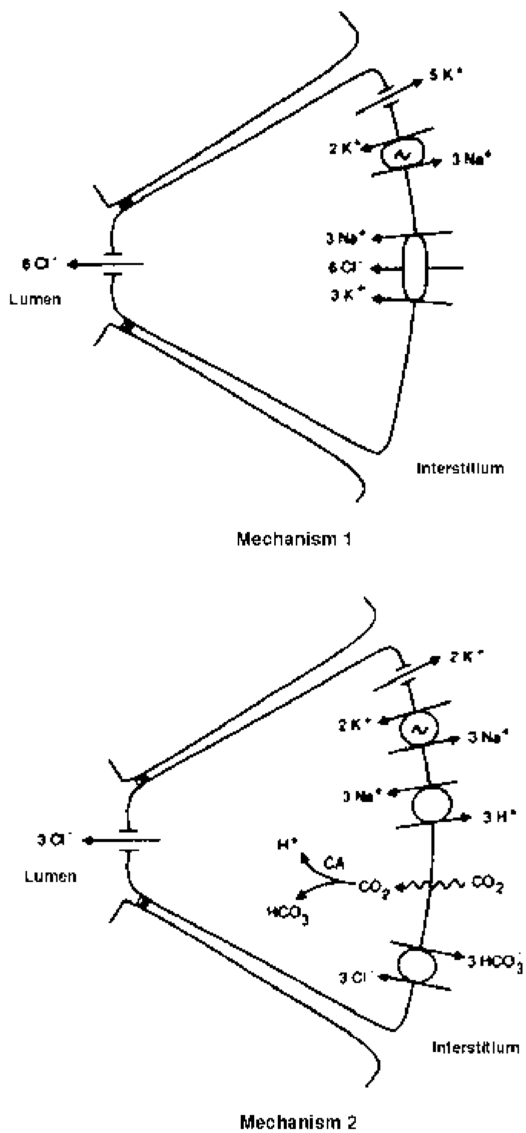
In contrast, acinar cell stimulation from the parasympathetic nervous system leads to high levels of fluid secretion (**Fig. 53–4A**). Acetylcholine receptors are triggered on the acinar cell membrane. These stimulate a G protein that subsequently activates phospholipase C (PLC). PLC then hydrolyzes phosphatidylinositol 4,5-bisphosphate to form two second messengers, inositol triphosphate (IP_3) and diacylglycerol (DAG). DAG promotes activation of protein kinase C and subsequent stimulation of a minor exocytic pathway and protein release. IP_3, however, plays a much more crucial role in calcium mobilization and in salivary gland fluid secretion.

IP_3 binds to a receptor that is located on an intracellular storage of calcium, most likely a portion of the endoplasmic reticulum (ER). After IP_3 binds to the ER, calcium ions are released down a concentration gradient that leads to a rapid 10-fold increase in intracellular calcium concentration within 5 seconds of acetylcholine receptor stimulation. The calcium increase then stimulates the opening of a chloride channel on the apical (luminal) side of the acinar cell.

An understanding of the effects of calcium on leading to fluid secretion by the acinar cell requires an explanation of the electrolyte composition and ionic transport systems of the acinar cell. The resting acinar cell contains a high concentration of K^+ and Cl^-. The K^+ ionic gradient is maintained by a Na^+/K^+ adenosinetriphosphatase



A



B

Figure 53–4 ***(A)*** *Mechanism of adrenergic and cholinergic stimulation of a salivary acinar cell. α, α-adrenergic receptor; ATP, adenosine triphosphate; β, β-adrenergic receptor; cAMP, cyclic adenosine monophosphate; DAG, diacylglycerol; Gp, G protein that stimulates phospholipase C (PLC); Gs, G protein that stimulates adenylyl cyclase; IP_3, inositol triphosphate; MUSC, muscarinic cholinergic receptor; PEPT, substance P receptor; PIP_2, phosphatidylinositol 4,5-bisphosphate; VIP, vasoactive intestinal polypeptide receptor. (From Baum BJ. Principles of saliva secretion. Ann NY Acad Sci 1993; 694:17–23. Reprinted with permission.)* ***(B)*** *Two mechanisms of electrolyte flux and replenishment during salivary acinar fluid secretion. (Adapted from Turner NJ. Mechanisms of fluid secretion by salivary glands. Ann NY Acad Sci 1993; 694:24–35. Reprinted with permission.)*

(ATPase) that exchanges Na^+ for K^+ on the basolateral side of the acinar cell. As the K^+ is imported into the cell, Na^+ is exported. Cl^- is concentrated in the cell by either a Na^+-K^+-$2Cl^-$ cotransportor or a Cl^-HCO_3 exchanger.

The rapid increase in intracellular calcium stimulated by IP_3 leads the opening of two important channels: a basolateral K^+ channel, and an apical Cl^- channel. K^+ and Cl^- then flow out of the acinar cell down their gradients, leading to the accumulation of Cl^- with an associated negative charge at the apical or luminal side of the cell (**Fig. 53–4B**). Na^+ follows the luminal Cl^- by electrical attraction, leaking from the interstitium through the tight junctions between acinar cells. The resulting osmotic gradient into the lumen caused by the flow of Na^+ and Cl^- results in a transepithelial movement of water from interstitium to lumen. The secretion of water into the acinar lumen results in an approximate 25% reduction of the acinar cell's volume within 10 to 15 seconds after maximal acetylcholine stimulation of the muscarinic receptors.

The flow of water into the acinar lumen is sustained by the continued entry of K^+ and Cl^- into the cell by a basolateral $Na^-K^{-2}Cl$ cotransporter to replenish the calcium-stimulated depletion K^+ and Cl^- (**Fig. 53–4B**). Cl^- may also be replenished within the acinar cells by a second mechanism. Carbonic anhydrase stimulates the conversion of CO_2 and H_2O to H^+ ion and HCO_3-. A basolateral Cl^-/HCO_3- exchanger may then import Cl^- into the cell and pump HCO_3- out of the cell. The remaining H^+ may then be removed from the cell by a Na^+/H^+ exchanger, using the gradient generated by the Na^+/K^+ ATPase already described. This removal of the H^+ ion through this mechanism decreases intracellular acidification created by bicarbonate loss from the Cl^-/HCO_3^- exchanger. The cellular pH has been measured to drop 0.1 unit as a result of the bicarbonate loss.

Ductal Modification and Transport

Although the acinar cells are water permeable, the ductal cells are water impermeable. The ductal system is composed of intercalated ducts, striated ducts, and interlobular ducts. As the salivary fluid passes through the ductal system, its electrolyte composition is modified. The striated ducts histologically resemble the distal tubules of the kidney and are metabolically active with a rich supply of enzymes, blood supply, and autonomic innervation. As the fluid passes through the striated ducts, most of the Na^+ and Cl^- are reabsorbed and a small amount of K^+, HCO_3^-, and proteins are secreted. Although the primary saliva produced in the acinus is isotonic with plasma, by the time the final saliva enters the mouth it is markedly hypotonic at ~25 mEq/L. The electrolyte modification of saliva is, in part, dependent on the salivary flow. Whereas rapid flow rates allow little time for ionic modification, slower flow rates permit an increasingly hypotonic saliva production. In contrast to the metabolically active striated ducts, the interlobular ducts function mainly for fluid transport. The interlobular ductal epithelium has collagen and elastic fibers on the external surface to allow passive stretching to accommodate varying volumes of passing saliva.

Salivary Flow Rates

Over a 24-hour period, the submandibular glands have been estimated to produce 71% of the total saliva volume, the parotid 25%, the sublingual gland 3%, and the minor salivary glands trace amounts. These proportions change depending on the amount of stimulation of the salivary glands. Under maximum salivary stimulation, the parotid gland's contribution may increase to 66% of the total volume.

The flow rate of stimulated saliva considerably exceeds that of baseline resting secretion. The unstimulated salivary gland produces anywhere from 0.001 to 0.2 mL per minute per gland. This flow may increase to 0.2 to 1.7 mL per minute per gland when stimulated. The 24-hour volume of total salivary secretion has been estimated to be 1000 to 1500 mL, although others have suggested that 500 to 600 mL is more realistic.

Salivary Components

Saliva contains various macromolecules that are produced by the salivary acinar cells. As described previously, the acinar cells release their protein products through exocytosis after stimulation by the sympathetic nervous. Although ~99.5% of saliva is water, there are a wide variety of substances in salivary fluid. In the parotid gland, salivary amylase is the protein predominantly secreted; in the submandibular gland, many mucinous glycoproteins are produced (see later discussion). The salivary glands contribute ~40% of all amylase secreted, with the remaining 60% secreted by the pancreas. Salivary amylase assists in polysaccharide digestion; however, one product produced by amylase, maltose, may be used by bacteria to produce acid. These acid products may lead to subsequent tooth enamel decay.

Salivary components including lysozyme, peroxidase, amylase, immunoglobulin A (IgA), and mucins play an important defensive role in the oral cavity. Lysozyme is present in large amounts; it hydrolyzes the mucopolysaccharide and mucopolypeptide components within the cell walls of gram-positive bacteria. Peroxidase is an enzyme that oxidizes substrates with hydrogen peroxide. The oxidation product of thiocyanate (SCN^-) is hypothiocyanate ($OSCN^-$), a substance that has been shown to have bacteriostatic activity. Amylase not only serves as a digestive enzyme but also may degrade the polysaccharide in the cell walls of gonococci.

Secretory IgA is important in the defense of superficial mucosal viral diseases. IgA provides a local immunity to viruses by a joint action between antibody binding and phagocytosis. Bacteriolysis by salivary IgA has not been demonstrated, although IgA does cause the aggregation of bacteria and facilitates the phagocytosis of bacteria. IgA is produced by plasma cells in the connective tissues of all salivary glands and is bound in a dimeric form by a “secretory component” protein produced by the salivary glandular cells. Salivary mucins may form complexes with IgA and lysozymes that permit a concentration of these antimicrobials within the oral cavity, enhancing their effectiveness. Additional defensive qualities of mucins include the provision of a protective coating and assistance in the mechanical clearance of bacteria.

Other salivary components include lactoferrin, albumin, globulin, cystatins, statherins, proline-rich proteins, and enzymes such as phosphatases, hydrolases, dehydrogenases, esterases, and carbonic anhydrases. An increasing number of polypeptides have been identified within saliva. These have included epidermal growth factor, nerve growth factor, renin, kalikreins, peptide hydrolases, and glucagon. These factors may have an endocrine or paracrine role function in homeostasis mechanisms, although their exact roles remain to be determined.

As described previously, saliva is hypotonic by the time it has passed the ductal system into the oral cavity. Parotid salivary electrolyte concentrations have been measured as follows: Na^+ 23 mEq/L, K^+ 20 mEq/L, Cl^- 23 mEq/L, HCO_3^- 20 mEq/L, Ca^+ 2 mEq/L, Mg^- 0.2 mEq/L, and HPO_4 6 mEq/L. Submandibular salivary electrolyte concentrations have been measured as follows: Na 21 mEq/L, K 17 mEq/L, Cl 20 mEq/L, HCO_3 18 mEq/L, and Ca 3.6 mEq/L, Mg 0.3 mEq/L, and HPO_4 4.5 mEq/L. In general, parotid gland electrolyte concentrations are slightly higher than those of the submandibular gland, with the exception of calcium, which is nearly twice as concentrated in the submandibular gland. This higher calcium concentration combined with a higher mucin content and a dependent inferior position makes the submandibular gland more likely to form sialoliths. Saliva has been shown to aid in enamel formation by providing a rich source of calcium, fluoride, phosphate, and magnesium. Bicarbonates serve to buffer saliva, maintaining a pH ranging from 5.8 to 7.1.

Salivary Mucins

The oral cavity is covered by a slimy, viscoelastic coat of mucus that forms a protective blanket on the oral mucosa. The viscosity of this layer is an important feature that permits lubrication of the oral cavity, facilitation of chewing and swallowing, protection of tooth minerals, and regulation of oral microbiological flora. This salivary viscosity is conferred by the O-glycosylated mucin-glycoproteins, or more simply termed mucins.

There are two distinct forms of salivary mucins. One type is composed of multiple bonded subunits and is large with high molecular weights. These mucins are termed MG_1 and are important in coating functions and lubrication during mastication and swallowing. In contrast, the other mucin type is smaller and consists of single polypeptide chains. These smaller mucins are termed MG_2 and exert their effects in solution via interactions with microorganisms.

Mucins are produced by the submandibular and sublingual glands and provide the viscosity of saliva. Salivary viscosity is directly related to the proportion of mucinous acinar cells in salivary glands. The relative salivary gland viscosities have been measured to be 13.4 for the sublingual gland, 3.4 for the submandibular gland, and 1.5 for the parotid gland. The predominantly serous parotid gland is the least viscous with its thin salivary product.

Microorganisms may attach to teeth and produce acids that lead to tooth demineralization. However, some protection is afforded by the salivary mucins, which adhere to the tooth surface and contribute to the formation of the enamel pellicle. This pellicle then serves as a barrier to the acids produced by microorganisms. MG_1 has a higher affinity for tooth mineral or hydroxyapatite than MG_2, and plays a more significant role in tooth protection. Studies have shown that tooth enamel pellicles formed from mucin-depleted saliva provides only 30% of the original protection observed with normal mucin-containing saliva. MG_2 may help to modulate the number and type of microorganisms that colonize the mouth by promoting the agglutination and clearance of certain organisms, while favoring attachment and proliferation of others.

Suggested Readings

Batsakis J. Salivary gland physiology. In: Cummings CC, ed. Otolaryngology—Head and Neck Surgery. St. Louis: Mosby; 1999:1210–1222

Baum BJ. Principles of saliva secretion. Ann NY Acad Sci 1993; 694:17–23

Garnett JR. Recent advances in physiology of salivary glands. Br Med Bull 1975;31:152

Graney DO, Jacobs JR, Kern RC. Salivary gland anatomy. In: Cummings CC, ed. Otololaryngology—Head and Neck Surgery. St. Louis: Mosby; 1999:1201–1209

Johns ME. The salivary glands: anatomy and embryology. Otol Clin North Am 1977;10:261–271

Kontis TC, Johns ME. Anatomy and physiology of the salivary glands. In: Bailey BJ, ed.: Head and Neck Surgery—Otolaryngology. Philadelphia: Lippincott-Raven; 1998:531–539

Mandel ID. Sialochemistry in diseases and clinical situations affecting the salivary glands. Crit Rev Clin Lab Sci 1980;12:321

Nauntofte B. Regulation of electrolyte and fluid secretion in salivary acinar cells. Am J Physiol 1992;263:G823–G837

Rice DH. Salivary gland physiology. Otol Clin North Am 1977;10: 273–285

Seifert G, Miehlke A, Haubrich J, et al. Diseases of the Salivary Glands. New York: Thieme Medical Publishers; 1986

Stuchell RN, Mandel ID. Salivary gland dysfunction and swallowing disorders. Otol Clin North Am 1988;21:649

Tabak LA. In defense of the oral cavity: structure, biosynthesis, and function of salivary mucins. Annu Rev Physiol 1995; 57:547–564

Turner JR. Mechanisms of fluid secretion by salivary glands. Ann NY Acad Sci 1993;694:24–35

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. Salivary flow is controlled by
   - A. The motor branches of the facial nerve
   - B. The parasympathetic nervous system
   - C. The sympathetic nervous system
   - D. A combination of sympathetic and parasympathetic systems

2. The bulk of total salivary flow volume is produced by
   - A. The submandibular glands
   - B. The minor salivary glands
   - C. The parotid gland
   - D. The sublingual gland

Chapter 54

MORPHOPHYSIOLOGY OF THE THYROID AND PARATHYROID GLANDS

CARL E. SILVER AND LANE KREVITT

THE THYROID GLAND

EMBRYOLOGY

The thyroid gland is the first of the endocrine derivatives of the pharynx to develop. On the 17th day of gestation, the thyroid gland originates from the foramen cecum that develops between the first and second pharyngeal arches and begins its descent ventrad to the pharynx as a patent bilobed diverticulum (i.e., the thyroglossal duct). By the third week, the thyroid gland reaches its destination at the midanterior neck. During migration the primoidum of the thyroid becomes associated with the ultimopharyngeal bodies of pharyngeal pouch V that detach from the pharynx and become incorporated into the dorsal aspect of the migrating thyroid. By the 10th week, the thyroglossal duct normally becomes fragmented and disappears, and the thyroid gland loses its connection to the pharynx. If this fails to happen, then ectopic thyroid tissue may be present just beneath the tongue or anywhere along the path of its descent. Remnants of the thyroglossal duct are a common finding in children, but the presistence of a patent thyroglossal duct is rare and when present could form a thyroglossal fistula that could enter into the median plane of the neck. In week 11, the characteristic morphology of the thyroid gland is present, and the gland is capable of concentrating iodides to synthesize thyroid hormones.

ANATOMY

The thyroid gland is a bilobed, midline structure. Each thyroid lobe is attached to the trachea by a dense consolidation of connective tissue, called the lateral suspensory (Berry's) ligament. Blood supply is by the superior thyroid branch of the external carotid artery and the inferior thyroid branch of the thyrocervical trunk. Superior and middle thyroid veins drain directly into the anterior and internal jugular veins. The inferior and thyroid ima veins are tributaries of the innominate veins. The recurrent laryngeal nerves are situated in close proximity to the thyroid and parathyroid glands, and may run superficial or deep to the inferior thyroid arteries. Lymphatic drainage of the isthmus and inferior lateral lobes is to the paratracheal and lower deep cervical nodes. The superior portions drain into the superior pretracheal and cervical nodes.

Physiology

The thyroid gland incorporates iodine into the amino acid tyrosine to form tetraiodothyronine, also called thyroxine (T_4), and triiodothyronine (T_3). Iodine uptake is controlled by thyroid-stimulating hormone (TSH), produced by the pituitary gland. The thyroid hormones are stored as thryroglobulin, the main component of colloid, and released into the circulation according to the influence of TSH. T_4 is thought to be a precursor to T_3, which exists in smaller quantities in the serum and is the active cellular hormone. Most of the circulating thyroxine is bound by protein, predominantly thyroxine-binding globulin. The small fraction of hormone not bound to protein is called the free fraction and is the fraction available for conversion to T_3.

Various proteins and hormone levels can be measured that may prove clinically useful. T_4 and T_3 may be directly measured, and in the proper context are accurate determinants of the level of thyroid function, although measurement of the free T_4 provides the most reliable index of thyroid function. As thyroid hormone levels increase, production of TSH by the pituitary diminishes, thus maintaining a physiological euthyroid status. TSH production increases when thyroid hormone levels decrease. Thus TSH levels are decreased in hyperthyroid states, and usually increased in hypothyroidism. In Graves' disease, a hyperthyroid state, TSH levels may be increased inappropriately, indicating that the basic defect may lie in the inability of the pituitary to regulate thyroid function by responding to increased thyroid hormone levels. Exogenously administered thyroid hormone should decrease TSH levels, thus maintaining a euthyroid state in patients who are being treated with the hormone. Employing thyroid hormone administration to suppress TSH levels is used to treat various thyroid tumors and other diseases. The effectiveness of suppression may be monitored by determination of TSH levels.

The incorporation of iodine into the thyroid gland to produce thyroglobulin can be measured by using various radioactive isotopes of iodine to measure radioiodine uptake. The overall level of uptake is a measurement of thyroid function. Scintigraphic images demonstrating distribution of the isotope within the gland are useful for determining the activity of tumors or nodules within the thyroid. Malignant tumors tend to concentrate radioiodine poorly, and this information is of some value in determining the indication for removal of thyroid nodules. Radioactive iodine can be employed therapeutically for treatment of overactive thyroid conditions, such as Graves' disease and toxic nodular goiter. Radioactive technetium behaves similarly to radioactive iodine and has also been employed for diagnostic scintigraphy.

Well-differentiated thyroid cancers produce thyroglobulin. After total thyroidectomy, thyroglobulin levels may decrease to insignificant amounts. A subsequent rise in thyroglobulin may indicate recurrence of the cancer. A variety of antibodies to thyroglobulin antigens are measurable in autoimmune (Hashimoto's) and lymphocytic thyroiditis. Medullary carcinomas of the thyroid, as well as the precursor C cell hyperplasia, produce excessive levels of the hormone calcitonin. This fact is useful for detecting the presence of medullary carcinoma, particularly in asymptomatic family members of a patient with the familial form of the disease, and is a valuable marker for recurrence of medullary carcinoma after treatment. Patients with multiple endocrine neoplasia syndrome of types IIA and IIB may have occult medullary carcinoma and should have calcitonin determinations to detect this condition.

Abnormal Physiology: Hyperthyroidism

Hyperthyroidism results from excess thyroid hormone in the peripheral tissues. Biochemical identification of hyperthyroidism is confirmed by elevation of T_3, T_4, and radioactive iodine uptake. Graves' disease (i.e., diffuse toxic goiter), an autoimmune process, and toxic nodular goiter comprise over 90% of the cases of hyperthyroidism. Graves' disease, the more common entity, usually occurs in the third or fourth decades of life. Females are affected seven times more often than males. The thyroid gland is diffusely enlarged and soft. In addition to the abnormal thyroid function tests, autoantibodies are detectable. Extrathyroidal manifestations of Graves' disease, which include exophthalmos, pretibial myxedema, and thyroid acropachy, help to differentiate it from other causes of hyperthyroidism. A patient may present with any of these signs before the onset of hyperthyroidism.

Toxic nodular goiters may be single or multiple. They occur more commonly in patients over age 50 and in female patients eight times more often than in males. Toxic nodules tend to function autonomously, suppressing the remaining gland. The cause of toxic nodules is unknown, but they are seen more frequently in areas of endemic goiter and iodine deficiency. In addition to the abnormally high circulating thyroid hormone levels, radioiodine scanning can reveal irregular areas of increased activity corresponding to the hyperfunctioning nodules.

Suppressive Therapy

Treatment of hyperthyroidism is suppressive or ablative. Suppressive therapy is administered with antithyroid drugs. This type of therapy is usually the first line of treatment in the attempt to restore euthyroid status. The treatment is nondestructive and allows for the possibility of remission. Thionamide drugs, such as propylthiouracil and tapazole, are the most commonly used antithyroid agents. These agents act to inhibit conversion of T_4 to T_3, or to prevent formation of T_3 and T_4. Generally, 4 to 6 weeks of treatment is required for full therapeutic effect. Many patients cannot tolerate these agents without serious drug reactions. Lithium has an antithyroid effect, although it is rarely used clinically for this purpose. It acts primarily to inhibit the release of preformed thyroid hormone. Patients with small toxic goiters, with only mild elevation of serum thyroid hormone levels, without associated clinical effects, or with toxicity that rapidly remits with medication may be treated with long-term suppression therapy without ablation.

Radioiodine Ablation

Ablation may be in the form of radioactive ablation or surgical ablation. Ablation is required for most adult patients who fail to develop prolonged remission after administration of suppressive therapy. Ablative therapy with radioactive iodine (iodine 131) is the most frequently used method in the United States for permanent control of hyperthyroidism due to Graves' disease. Although low-dose irradiation is a known etiologic factor in thyroid carcinoma, an increased incidence of thyroid or other malignancies in patients treated with radioiodine has not been demonstrated. In the case of solitary nodules, surgical excision may be a simpler, more rapid and safer form of therapy, especially if the entire contralateral lobe can be preserved.

Age is another consideration in the decision between iodine 131 and surgery. Clinicians differ in their opinion as to the earliest age at which iodine 131 is considered safe to employ; the range varies from 20 to 40 years of age. Some patients, particularly in the younger age groups, refuse radioactive therapy. Pregnancy is a contraindication, and the expectation of future pregnancies may also influence against employment of radioiodine treatment.

Euthyroid status must precede treatment with iodine 131, to avoid toxic effects after treatment. Most patients are cured with a single ablative dose: more than two doses are seldom required. Hypothyroidism is the major common complication. Other rare complications include thyroid neoplasm, leukemia, genetic defects, thyroid crisis, and hypoparathyroidism.

Surgery for Hyperthyroidism

Indications for surgery for hyperthyroidism include failure of nonsurgical management, massive goiter with or without local symptoms, florid disease, desire for rapid control of the toxic process, patient bias, pharmacological intolerance, and suspicion of malignancy.

Surgery should not be performed without preoperative preparation to establish euthyroid status, in an attempt to avoid "thyroid storm" postoperatively. A thyroid storm is a massive outpouring of thyroxine and adrenergic hormones that may prove fatal. Most regimens for preoperative preparation have employed thionamide antithyroid agents until normal thyroid function was restored. Iodine, typically in the form of Lugol's solution or a saturated solution of potassium iodide, has often been administered for 10 to 14 days before surgery. Iodine is thought to increase the amount of colloid within the gland, thereby enlarging it, decreasing its vascularity, and making it less friable. It does not restore euthyroid status and may fail to prevent thyroid storm if used as the sole preoperative agent. Although frequently employed in the past, the value of iodine for preparation for surgical treatment of hyperthyroidism is questionable.

Beta-adrenergic blocking agents have proven effective for rapid preoperative preparation of hyperthyroid patients. These agents permit safe performance of surgery in 24 to 72 hours. In addition to their benefit in emergency situations, there are fewer adverse reactions than with thionamide drugs, and the diminution of preoperative delay and monitoring may be cost effective. Beta-blockers work by blocking the sympathetic manifestations of thyrotoxicosis. They do not alter all the metabolic derangements and are not appropriate for long-term management of hyperthyroidism.

A question remains as to whether, in nonemergency situations, it is preferable to employ beta-blockers alone, or to first control hyperthyroidism with antithyroid drugs and use beta-blockers only if the clinically hyperthyroid state cannot be brought under control within a reasonable time before surgery. At present, antithyroid drug preparation, with or without iodine or beta-blocking drugs, appears to be the safest method of preoperative preparation. If complete preoperative control of hyperthyroidism is not feasible because of drug intolerance, lack of compliance, or other factors, rapid preoperative preparation with beta-blocking agents will permit safe performance of surgery.

Graves' disease is treated surgically by subtotal thyroidectomy, usually bilateral. Unilateral toxic nodular goiter is treated by lobectomy, and bilateral nodular goiter is treated by subtotal thyroidectomy. Although malignancy is rarely associated with hyperthyroidism, all nodules found within a hyperthyroid gland should be evaluated in a fashion similar to those in a euthyroid gland to rule out malignancy.

Surgical treatment consists of removal of sufficient hyperfunctioning thyroid tissue to reduce circulating thyroid hormone levels to normal or hypothyroid levels. Bilateral subtotal thyroidectomy has been the traditional method of treatment, although total, or "near total" thyroidectomy is often employed to prevent recurrence or to treat multiple toxic nodules. The complication rate of surgery is low. The most common complication reported is transient hypocalcemia. Permanent hypocalcemia is rare. Vocal cord paralysis occurs in less than 2% of patients. Postoperative clinical or biochemical hypothyroidism occurs in ~25% of patients treated with subtotal thyroidectomy. This condition is readily managed by thyroid hormone replacement and is preferable to hyperthyroidism as a result of inadequate resection. The incidence of post-treatment hypothyroidism is lower after surgical treatment than irradiation, but both are associated with a significant incidence of long-term hypothyroidism. Recurrent hyperthyroidism reflects failure of the procedure.

THE PARATHYROID GLANDS

Embryology

The inferior parathyroid glands develop from the third arch pharyngeal pouch, and the superior parathyroid glands originate from the tissue of the fourth arch pharyngeal pouch. One of the tissues that contribute to the formation of these glands is thought to be neuroectoderm (neural crest tissue) even though these glands are derived from the pharyngeal pouches; therefore, the parathyroid glands probably contain contributions from two germ layers (i.e., endoderm and ectoderm) and neural crest cells. Fetal parathyroid glands are functional, being responsive to calcium levels in the blood and producing parathormone that has been detected in fetal blood. Only the chief cells are present in the fetal parathyroid glands, and the other cell types (e.g., oxyphils) differentiate postnatally.

Clinical Considerations

The otolaryngologist is concerned with parathyroid physiology in its pathological hyperactive state. Hyperparathyroidism, either from neoplasm or diffuse hyperplasia, is treated surgically. The discussion will be limited to hyperparathyroidism, associated anatomy, pathological considerations, and surgical management.

Once considered a rare disease, hyperparathyroidism is now encountered with considerable frequency by the head and neck surgeon. The reason for the apparent increase in incidence stems from several factors, including availability of an accurate serum calcium determination by modern automated laboratory techniques, the availability of parathyroid hormone assay to confirm the diagnosis of hyperparathyroidism, and the appearance of a group of patients with secondary and tertiary hyperparathyroidism as a result of prolonged maintenance on renal dialysis.

Patients with hyperparathyroidism are generally seen within several clinical contexts. The most common is sporadic (nonfamilial) primary hyperparathyroidism. Far less frequently encountered are patients with familial hyperparathyroidism, or multiple endocrine neoplasia (MEN) syndromes. Renal dialysis patients with parathyroid hyperplasia form a third group that represents secondary and tertiary hyperparathyroidism.

Surgical Anatomy

Superior and inferior parathyroid glands are each situated within a range of anatomical locations, based on their embryologic derivations; the superior from the fourth branchial pouch and the inferior from the third. Thus the superior gland tends to be posterior (or dorsal) in location and is situated posterior to the recurrent laryngeal nerve, most often on the posterolateral surface of the thyroid lobe. When it migrates, it tends to descend in the prevertebral plane toward the posterior mediastinum. It is often paraesophageal and occasionally retroesophageal in location.

The inferior parathyroid glands are derived from the third branchial pouch and descend in conjunction with the thymus gland. The inferior parathyroid glands are more variable in location than the superior glands, due to the longer path of descent. The glands are situated further anteriorly (ventral) than the superior glands and lie anterior to the recurrent laryngeal nerves, most often at or near the lower poles of the thyroid gland. The inferior parathyroid gland is often situated within the thyrothymic ligament and may descend into all portions of the thymus gland, including the portions located within the deep anterior mediastinum. Occasionally, the inferior gland fails to descend and may be situated near the superior pole of the thyroid. Occasionally, such lesions are found in the carotid sheath at the level of the hyoid bone.

It is important when conducting parathyroid exploration to document the location of each parathyroid gland that is found. Thus the surgeon is aware of which parathyroid glands have not been found and may be guided as to where to search for the missing gland.

Supernumerary glands occur in ~5% of patients and may present a problem with regard to finding a parathyroid lesion or in curing a patient after an apparently complete exploration has been performed. Supernumerary glands are often ectopic in location, thus compounding the problem. The availability of accurate preoperative imaging studies and intraoperative parathyroid hormone determination has helped in dealing with the problems presented by supernumerary and ectopic parathyroid lesions.

Pathology Considerations

Parathyroid adenoma, the most common cause of hyperparathyroidism, is a benign neoplastic condition involving usually only one gland, distinguished by a rim of compressed normal tissue located within the capsule of the lesion. Parathyroid hyperplasia is defined as generalized abnormality and potential hyperfunction of all parathyroid tissue. Although many authorities believe that a clear-cut distinction exists between adenoma and hyperplasia in sporadic primary hyperparathyroidism, clinical and pathological evidence may be equivocal. Our experience has demonstrated that the distinction between adenoma, multiple adenomas, and hyperplasia is difficult to make, except in cases of secondary or tertiary hyperparathyroidism associated with renal disease and in familial disease. In primary hyperparathyroidism, differentiating adenoma from hyperplasia is difficult because microscopic abnormalities are similar in both entities and microscopically abnormal glands may be grossly normal in size and appearance. Hypercellularity, diminished fat, microscopic nodular hyperplasia, abnormal cytology, and oxyphilic nodules may exist in both conditions. The clinical implication for this distinction is especially important with regard to the now popular avocation for unilateral parathyroid exploration, based on preoperative imaging studies, with confirmation of the diagnosis of adenoma, by finding of an abnormal as well as a normal gland on the side explored. This is possible in many cases because the majority of patients with primary hyperparathyroidism have a single gland, grossly enlarged and microscopically abnormal, with the remaining three glands grossly and microscopically normal. The recent availability of rapid intraoperative parathyroid hormone determination tissue has simplified the problem and informs the surgeon whether all hyperactive parathyroid tissue has been removed at any given point during the operation.

Diagnosis: Parathyroid Hormone and Hypercalcemia

It is essential for the surgeon to be certain that the diagnosis of hyperparathyroidism is correct. Virtually all patients with primary hyperparathyroidism have hypercalcemia. Ionized calcium levels may provide a more accurate appraisal of the patient's abnormal calcium metabolism than the more frequently performed total calcium measurements. Serum phosphorous level is below normal in approximately one third of the patients and at the lower range of normal (2.3–3.0 mg/dL) in most others. Other laboratory values may not be abnormal and are not consistent indicators of the diagnosis.

The most important documentation of hyperparathyroidism is the finding of elevated levels of parathyroid hormone (PTH) in the serum. PTH levels are suppressed in almost all conditions that produce hypercalcemia. Previously, fragment determinations by radioimmunoassay for the circulating carboxy terminal (C terminal) or midregion portions of the molecule were performed with reasonable reliability. At present, two site assays by immunoradiometric (IRAM) or immunochemiluminometric (ICMA) methods for intact PTH are preferred. Eight-five to 90% of patients with primary hyperparathyroidism have elevated levels of circulating intact PTH. The others have levels in the upper range of normal, which is inappropriate in view of the tendency for PTH levels to be suppressed in nonhyperparathyroid hypercalcemia.

The most common cause of nonparathyroid hypercalcemia is malignancy. Many tumors produce a parathyroid hormone–related protein (PTHrP), which causes a syndrome similar to hyperparathyroidism and paraneoplastic syndromes. In rare instances, malignant tumors of nonparathyroid origin may produce intact PTH.

Other causes of hypercalcemia account for only 10% of cases. The numerous causes of hypercalcemia are listed in **Table 54–1**.

TABLE 54–1 DIFFERENTIAL DIAGNOSIS OF HYPERCALCEMIA

Primary hyperparathyroidism
Sporadic (adenoma, hyperplasia, carcinoma)
Familial
Isolated
Cystic
Multiple endocrine neoplasia, type I or II
Malignancy
PTHrP
Excess production of 1,25-dihydroxyvitamin D
Other factors (cytokine, growth factors)
Disorders of vitamin D
Exogenous vitamin D toxicity
Endogenous production of 25-hydroxyvitamin D (Williams syndrome)
Endogenous production of 1,25-dihydroxyvitamin D
Granulomatous disease
Sarcoidosis
Tuberculosis
Histoplasmosis
Coccidiomycosis
Leprosy
Others
Lymphoma
Nonparathyroid endocrine disorders
Thyrotoxicosis
Pheochromocytoma
Acute adrenal insufficiency
Vasointestinal polypeptide hormone–producing tumor
Medications
Thiazide diuretics
Lithium
Estrogens/antiestrogens, testosterone in breast cancer
Milk-alkali syndrome
Vitamin A toxicity
Familial hypocalciuric hypercalcemia
Immobilization
Parenteral nutrition
Aluminum excess
Acute and chronic renal disease

PTHrP, parathyroid hormone–related protein.
Adapted from Chan FKW, Koberle LMC, Thys-Jacobs S, Bilezikian JP. Differential diagnosis, causes, and management of hypercalcemia. Curr Probl Surg 1997;34:445–523.

Indications for Treatment

Not all patients with hyperparathyroidism producing asymptomatic hypercalcemia require operation. The following criteria are now commonly accepted:

- Serum calcium >12 mg/dL hypercalcemia >400 mg/24 hours
- Any overt manifestation of primary hyperparathyroidism
- Reduced creatinine clearance in the absence of other causes
- Age <50

Many "asymptomatic" patients may have long-standing symptoms to which they have become acclimated. After surgery, many such patients regain a level of mental alertness, physical stamina, and relief from chronic bone and joint aches that represent a significant improvement over their preoperative status. Careful history taking and a comprehensive laboratory analysis, including bone density studies, are essential before rejecting a patient with well-documented hyperparathyroidism for treatment.

The indications for surgery in patients with secondary, or renal, hyperparathyroidism are complex and beyond the scope of this discussion. In general, however, surgery is recommended for these patients because of painful and debilitating bone disease (renal osteodystrophy). In recent years, patients with massive elevations of intact PTH have been referred for surgery prior to developing severe osteodystrophy to prevent the severe effects of this condition.

Localization Studies

Parathyroid surgery has always been challenging because of the variability in the locations of parathyroid glands and the difficulty of demonstrating the lesions with conventional imaging studies. Except for rare parathyroid cancers, enlarged parathyroid glands are almost never palpable on physical examination.

Recently, considerable interest has centered around the use of technetium 99m labeled sestamibi, used either alone or in conjunction with iodine 123 technetium 99m pertechnate subration scanning. The single radionuclide method of sestamibi parathyroid imaging is based on the differential rate of "washout" of the sestamibi from thyroid and parathyroid tissue. Because the parathyroid glands retain the substance longer than the thyroid gland, an enlarged and hyperfunctioning parathyroid gland will be displayed on the delayed image. A useful adjunct to the sestamibi imaging technique is the use of single photon emission computed tomography (SPECT), with reconstruction in the transaxial, coronal, and sagittal planes. Eighty percent or more of solitary parathyroid lesions can be demonstrated by noninvasive sestamibi scanning and its adjuvant procedures. Preoperative scanning is less useful for localization of all lesions in cases of multiple gland disease, but it is particularly helpful in identification of ectopic lesions, which can always be present in either solitary or multiple gland disease. These lesions are often found in intrathroacic or upper cervical locations. Other studies, such as ultrasound, magnetic resonance imaging, and computed tomographic scans, may demonstrate parathyroid lesions in various cases, but none of them have been as useful as scintigraphy. Invasive procedures, such as selected vein catheterization and arteriography, may be required for more definitive preoperative localization, especially in revision surgery.

Suggested Readings

Burman KD. Thyroid Cancer II. Endocrinol Met Clin North Am 1996;25:1–211

Callender DL, Sherman SI, Gagel RF, Burgess MA, Geopfert H. Cancer of the thyroid. In: Myers EN, Suen JY, eds. Cancer of the Head and Neck. 3rd ed. Philadelphia: WB Saunders; 1996;485–515

Chan FKW, Koberle LMC, Thys-Jacobs S, Bilezikian J. Differential diagnosis, causes, and management of hypercalcemia. Curr Probl Surg 1997;34:445–523

Kaplan MN, Larsen PR. Thyroid disease. Med Clin North Am 1985;69:847

Kaplan EL, Yoshiro T, Salti G. Primary hyperparathyroidism in the 1990s. Ann Surg 1992;215:300–317

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. Which serologic marker may be indicative of recurrent well-differentiated thyroid carcinoma?
   - A. Decreased thyroglobulin
   - B. Increased thyroglobulin
   - C. Increased thyrotropin
   - D. Increased thyroid-stimulating hormone (TSH)
   - E. Decreased TSH
2. The thyroglossal duct
   - A. Is a normal structure in the neck region of newborn infants
   - B. Represents the anlagen of the thyroid glands and normally disappears by week 10 of gestation
   - C. Is persistent in the midneck region of most young adults
   - D. Contributes to the formation of the ultimopharyngeal bodies
   - E. Is formed by the dorsal division of the first pharyngeal pouch
3. What is the most common cause of hypercalcemia?
   - A. Parathyroid adenoma
   - B. Malignancy
   - C. Idiopathic
   - D. Hyperplastic thyroid disease
   - E. Familial hypocaluric hypercalcemia
4. Evidence of hypercellularity is diagnostic for parathyroid carcinoma.
   - A. True
   - B. False
5. In which instance(s) is/are thyroid carcinoma associated with parathyroid disease?
   - A. Multiple endocrine neoplasia type I (MEN I) and MEN IIA
   - B. MEN I and MEN IIB
   - C. MEN IIA and MEN IIB
   - D. MEN I
   - E. All of the above

Chapter 55

PATHOBIOLOGY OF THE THYROID GLAND

Margaret Brandwein-Gensler

BENIGN LESIONS

Goiter

Goiter (from the Latin *guttur*, meaning "throat") is a clinical term used to describe any thyroid enlargement. The term generally refers to enlargement resulting from a variety of benign conditions: physiological (puberty, pregnancy), metabolic (dietary iodine deficiency, i.e., endemic goiter), abnormal iodine metabolism, or inflammatory/autoimmune diseases (Hashimoto's disease, Graves' disease). Because this term lacks specificity, thyroid malignancies also appear as a thyroid enlargement or mass. Early goiters may present as simple, smooth, or mildly enlarged, palpable glands. Prolonged thyroid enlargement results in the formation of multiple adenomatous nodules, which may undergo hemorrhage and fibrosis, leading to a multinodular (adenomatous) goiter. Although adenomatous goiters are usually benign, they can cause serious clinical problems. The gland can be grossly disfiguring, encroach upon the trachea, or produce respiratory obstruction or hoarseness from recurrent laryngeal nerve paresis. Intrathoracic goiters can obstruct venous return (superior vena cava syndrome) and cause severe respiratory difficulties. As goiters increase in size, they can impinge and distort the trachea or esophagus, resulting in dyspnea and dysphagia.

On gross examination, a multinodular goiter may present as small nodules distributed diffusely or asymmetrically throughout the gland. Cut sections reveal multiple nodules, their glassy, pink/mahogany surface reflecting colloid storage. Histologically, one sees multiple follicles with flattened thyrocytes and colloid accumulation, with fibrosis, hemorrhage, calcification, and even ossification.

Adenomas

Follicular Adenoma

The most common solitary thyroid mass is an adenoma. This is an encapsulated, benign expansion of thyroid follicles, compressing adjacent thyroid parenchyma. On gross inspection, the capsule is usually thin and flimsy; a thickened capsule can raise the diagnostic possibility of a minimally invasive follicular carcinoma (this term is defined later). Thyroid follicles can be large and distended by colloid, or smaller with variable colloid content. Evidence of hyperactivity can be seen. If an adenoma is significantly hyperfunctional (e.g., toxic adenoma), then the histological distinction between it and an encapsulated papillary thyroid carcinoma can be difficult. This is because stimulated, hyperfunctioning thyroid will form hyperplastic pseudopapillae, as will be discussed later. Adenomas may have a varied follicular architecture, which is of no prognostic significance. However, some variants of follicular adenomas require special mention.

Hürthle Cell Adenoma

Hürthle cells refer to thyrocytes that have undergone oncocytic metaplasia, a process of unknown significance that can occur in many endocrine and nonendocrine organs; it occurs in various processes: inflammatory (Hashimoto's thyroiditis), benign (Hürthle cell adenoma), or malignant (Hürthle cell carcinoma). Macroscopically, a Hürthle cell adenoma is an encapsulated mass with a deep mahogany hue. This color can be seen in any histologically oncocytic neoplasm (renal cell oncocytoma, parotid oncocytoma, etc.). Microscopically, the term *oncocyte* refers to cells with abundant, bright pink, granular cytoplasm due to increased mitochondria. Hürthle cells are large thyrocytes that have undergone oncocytic change. The cells can also be pleomorphic (i.e., atypical) and therefore worrisome. However, the criteria for malignancy remains the same for all follicular neoplasia: in the absence of capsular invasion or subcapsular vascular invasion, all follicular lesions, including all Hürthle cell adenomas, regardless of pleomorphism, are still benign adenomas. The distinction between Hürthle cell adenomas and carcinomas (as well as follicular adenomas and minimally invasive follicular carcinomas) relies on architecture (capsular or vascular invasion), not cytological features, such as cellular pleomorphism or mitoses. Neither fine-needle aspiration nor frozen section examination can address this distinction. Fine-needle aspiration permits only cytological evaluation. Frozen section examination allows only limited tissue sampling as compared with permanent section evaluation; focal diagnostic findings of invasion may not be observed. The size of Hürthle cell tumors may be helpful because it correlates strongly ($p < .005$) with the likelihood of malignancy. Large tumors (>4 cm) are more likely to be malignant than small tumors (<2 cm).

Atypical Adenoma

Historically, the term *atypical adenomas* was applied to follicular adenomas (especially Hürthle cell adenomas) that possessed atypical features (nuclear atypia, necrosis) but fell short of the criteria required for malignancy (capsular invasion or subcapsular vascular invasion). They were often reported as "indeterminate" or "possibly malignant." Importantly, follow-up studies have shown that tumors classified as "indeterminate" behave in a clinically benign fashion, removing the need for this term.

Hyalinizing Trabecular Tumor

This uncommon subtype (also referred to as benign hyalinizing trabecular adenoma, or malignant hyalinizing trabecular carcinoma) may be confused with thyroid malignancies; its own oncologic potential has been questioned. It appears as an encapsulated tan mass, microscopically composed of polygonal, oval, and elongated tumor cells arranged in cords (trabeculae) or clusters (zell ballen) with minimal follicle formation. The cellular cytoplasm of the tumor is finely granular, oncocytic, or optically clear. Similar to papillary thyroid carcinoma, its nuclei are oval, elongated, with inclusions and grooves. Rare psammoma bodies also can be seen. The differential diagnosis includes medullary carcinoma and papillary thyroid carcinoma. Immunohistochemistry of hyalinizing trabecular adenoma confirms expression of thyroglobulin, but not calcitonin, distinguishing it from medullary carcinoma. The distinction between hyalinizing trabecular adenoma and papillary thyroid carcinoma can be difficult, and in fact bona fide papillary thyroid carcinoma can develop within hyalinizing trabecular adenoma, raising suspicions that hyalinizing trabecular adenoma may be a precursor of papillary thyroid carcinoma. To wit, *RET/PTC* gene rearrangements are common to papillary thyroid carcinoma and have been detected in hyalinizing trabecular adenoma. To date, tumors reported as typical hyalinizing trabecular adenomas have followed benign courses.

Black Thyroid

Black thyroid is a curious incidental finding caused by minocycline ingestion, a tetracycline derivative administered in the treatment of acne. Patients on long-term, high-dose minocycline therapy may develop pigment deposition in numerous organ systems including the thyroid. This is thought to result from an oxidative interaction between minocycline and thyroid peroxidase; it may remain even after cessation of the drug. Grossly, the thyroid gland is black, and microscopically, the thyrocytes and colloid are pigmented. The pigmentation itself has no adverse affect on the thyroid gland. The incidental finding of thyroid pigmentation is usually limited to the surrounding gland and spares the adenoma or carcinoma that prompted surgery.

THYROIDITIS

Hashimoto's Thyroiditis

Diagnosis and Pathobiology

Hashimoto's thyroiditis is an autoimmune thyroiditis characterized by lymphocytic infiltration and destruction of thyroid with eventual loss of function. There is a marked female predominance; it is usually diagnosed during or after the fourth decade of life. Patients are usually euthyroid, but they may also present with hypothyroidism or hyperthyroidism (Hashitoxicosis). The usual presentation is that of a painless, firm thyroid mass. The gland may be asymmetrically enlarged, clinically simulating malignancy. Surgeons considering the diagnosis of Hashimoto's thyroiditis in these circumstances may avoid unnecessary thyroidectomies. The inflammation and fibrosis that often accompany Hashimoto's thyroiditis make surgical dissection especially difficult, with increased risk of damage to the recurrent laryngeal nerves and parathyroid glands. Fine-needle aspiration and serological studies (see later discussion) can lead to the diagnosis preoperatively. Serologically, patients with Hashimoto's thyroiditis have elevated antithyroid antibodies, mainly antithyroperoxidase, antithyroglobulin, and antimicrosomal antibodies. Antimicrosomal antibodies are specific but not sensitive in establishing the diagnosis. Cytological examination reveals abundant lymphocytes and Hürthle cells; fine-needle aspiration may have a good sensitivity to detect a diffuse process like Hashimoto's thyroiditis, but sampling limitations may fail to detect concomitant malignancy.

Hashimoto's thyroiditis is caused by abnormal activation of helper T lymphocytes within the thyroid. This may occur after viral infection or as a result of aberrant thyrocyte expression of major histocompatibility (MHC) class II proteins; these are the membrane glycoproteins responsible for immune distinction between "self" and "nonself." The activated T cells then recruit B lymphocytes, which produce a variety of antithyroid antibodies, and cytotoxic T cells that destroy thyrocytes. Thyroids affected by Hashimoto's thyroiditis are usually tan in color, reflecting the lymphocytic infiltrate. Microscopically, lymphocytes are seen infiltrating thyroid and forming germinal centers. The thyrocytes may undergo oncocytic metaplasia (Hürthle cell change).

Eventually patients with Hashimoto's thyroiditis have depleted thyroid function, but they may remain euthyroid or develop episodic, fluctuating thyroid dysfunction. Less common complications include episodic hyperthyroidism, or Hashimoto's encephalopathy. These patients are at risk for developing other autoimmune conditions, as well as thyroid malignancies (see next section). Hashimoto's thyroiditis is managed by monitoring thyroid function and by providing synthroid replacement for hypothyroidism.

Incidence of Neoplasia in Hashimoto's Thyroiditis

Lymphoma

There is a clear association between Hashimoto's thyroiditis and increased risk of thyroid lymphoma. The vast majority of thyroid lymphomas are encountered in patients with Hashimoto's thyroiditis. The usual presentation is that of a dramatically enlarging thyroid mass in an older woman, mimicking anaplastic carcinoma. The diagnosis may be established by fine-needle aspiration or core needle biopsy, avoiding unnecessary thyroidectomy in these cases. Histologically, most of these lymphomas are of B-cell origin and are usually a diffuse large B-cell lymphoma. Immunohistochemistry is routinely performed on such cases to make the distinction between lymphoma and other malignancies, (e.g., anaplastic carcinoma, undifferentiated squamous cell carcinoma) and then to further type the lymphoma. Many older literature reports of survivors of anaplastic thyroid carcinoma erroneously contained patients with large-cell malignant lymphoma, falsely improving the survival statistics.

Fine-needle aspiration has the potential to establish the diagnosis preoperatively. The next diagnostic option includes core needle or open biopsy, with frozen section analysis. This allows for fresh tissue to be saved for lymphoma markers. Only general lymphocyte, B-cell, and T-cell markers can be immunohistochemically studied using formalin-fixed paraffin-embedded tissue. The diagnostic sensitivity and specificity of fresh tissue studies are much greater: a vast array of lymphocyte markers can be detected, and B-cell clonality can be confirmed. Once a diagnosis of thyroid lymphoma is established, surgery is not a therapeutic option. Appropriate therapy is determined by clinical staging. The majority of cases present as either stage IE (confined to thyroid) or IIE disease (positive cervical and/or mediastinum lymph nodes). Radiotherapy is indicated for stage IE disease, whereas adjuvant chemotherapy is preferred for patients with stage II, III, and IV disease. Overall, thyroid lymphomas have a favorable outcome, but prognosis depends on clinical stage and histology: high-grade lymphomas and stage greater than IE have the greatest potential for a poor outcome. The overall relapse-free survival at 5 years is 72%, and the overall survival at 5 years is 88%.

Papillary Thyroid Carcinoma

The incidence of papillary thyroid carcinoma for patients with Hashimoto's thyroiditis is greater than expected, which initially suggested a causal relationship. But the burden of proving causality is a difficult task because both entities are relatively common, and their association may be considered merely coincidental. Lymphocytic infiltrate is common adjacent to papillary thyroid carcinoma and cannot be taken as evidence of Hashimoto's thyroiditis. In addition, cytologic atypia is common in Hashimoto's thyroiditis and may lead to the overdiagnosis of papillary thyroid carcinoma. Be that as it may, the *RET/PTC1* and *RET/PTC3* oncogenes, highly specific for thyroid malignancy, have been found in 95% of the Hashimoto's patients studied, without papillary thyroid carcinoma, suggesting a preneoplastic condition.

Graves' Disease

Diagnosis and Pathobiology

Graves' disease is an autoimmune thyroiditis resulting in hyperthyroidism. There is a female predominance of 6:1 to 10:1. Patients typically present with a goiter and hyperthyroidism. Radioactive iodine uptake is usually diffusely increased, but it may be localized to a dominant nodule (toxic adenoma). Ophthalmopathy (exophthalmos, impaired ocular motility, diplopia) is frequently present and can alert the clinician to consider the diagnosis of Graves' disease. It is the most frequent cause of unilateral exophthalmos. In contrast to Hashimoto's thyroiditis, in which the thyroid parenchyma constitutes the actual lymphocytic battleground, Graves' disease can be considered as an "assault from afar." Host lymphocytes produce an array of antibodies that stimulate the thyroid-stimulating hormone (TSH) receptor, resulting in hyperthyroidism. However, the lymphocytic infiltration and mediated thyrocyte destruction present in Hashimoto's thyroiditis are not seen in Graves' disease. As in Hashimoto's thyroiditis, thyrocytes of Graves' disease aberrantly express MHC antigens, thus stimulating immune reaction.

Graves' disease may be further complicated by pretibial myxedema, a nonpitting edema of the lower legs that invariably follows the hyperthyroidism and ophthalmopathy, and is probably a result of antithyroid antibodies cross-reacting and stimulating fibroblast receptors, resulting in glycosoaminoglycan overproduction. Graves' ophthalmopathy is due to swelling of the extraocular muscles and retrobulbar fat. Similar to pretibial edema, it is due to excessive production of glycosoaminoglycans by orbital fibroblasts. The ophthalmopathy is the result of a direct autoimmune targeting of orbital fibroblasts, probably through TSH receptor antibodies. After thyroid suppression or radioactive ablation, a Hashimoto's-like lymphocytic infiltrate remains; the aftermath of this battle is hypothyroidism, which will require synthroid replacement. The severe complications of Graves' ophthalmopathy (corneal abrasion, impending blindness) require orbital decompression.

Diagnostic Differentiation from Papillary Thyroid Carcinoma

A thyroid gland involved by Graves' disease is enlarged, reddened, and "meaty," histologically manifesting colloid depletion and follicular hyperplasia. If colloid is present, it is watery (nondense) or scalloped. The thyrocytes are tall and form pseudopapillae. A hyperfunctioning adenoma is an encapsulated lesion, with the identical hyperplastic papillary histological findings. The distinction between the follicular hyperplasia of Graves' disease and papillary thyroid carcinoma may at times be difficult, especially during frozen section. The nuclear morphology, which aids in the recognition of papillary thyroid carcinoma (optically cleared nuclei, nuclear holes, nuclear grooves), is more reliably observed in the routine formalin-fixed "permanent sections" than in the nonfixed frozen section slides. In papillary thyroid carcinoma, the tumor nuclei are enlarged, oval, and overlapping, whereas the nuclei of hyperplastic thyrocytes remain smaller and rounded.

Subacute Thyroiditis: De Quervain's (Giant Cell) Thyroiditis

De Quervain's thyroiditis (giant cell, granulomatous, or subacute granulomatous thyroiditis) is a transient thyroiditis. It is important to recognize this clinical entity, which is usually self-limiting, to avoid unnecessary surgery. Patients present with thyroid tenderness, referred ear pain, goiter, elevated serum T_3 or T_4, and elevated erythrocyte sedimentation rate. The thyroid is diffusely enlarged and firm. There is a history of antecedent upper respiratory infection, but patients also may present with an acute, febrile illness with weakness and malaise. De Quervain's thyroiditis is considered a nonautoimmune, postviral thyroiditis.

The diagnosis can be established by clinical history, physical examination, and fine-needle aspiration that reveals plump transformed thyrocytes, epithelioid granulomas, multinucleated giant cells, and an acute and chronic inflammatory background. Thyroid scintigraphy frequently reveals uniform decreased uptake, but at times, the decreased uptake can be localized to one lobe or to a "cold spot." Thyroid ultrasound will demonstrate a diffusely enlarged, hypoechogenic gland, and this modality can be utilized to monitor patient course. Surgery is contraindicated. The goal of therapy is symptomatic relief. Salicylates and nonsteroidal anti-inflammatory drugs are all effective therapies; corticosteroids may be necessary for more severe forms.

Riedel's Thyroiditis

Riedel's thyroiditis (Riedel's struma, invasive fibrous thyroiditis) is an extremely rare condition characterized by "invasive" fibrosis of the thyroid and a systemic inflammatory fibrosclerosing process. The usual presentation is that of a stone-hard thyroid mass, suggestive of malignancy. Patients may be dysphagic, dyspneic, hypothyroid, and even have vocal cord paralysis. Further systemic fibrosis may manifest as retroperitoneal fibrosis, sclerosing cholangitis, aortic sclerosis, renal cortical fibrosis, occlusive phlebitis, and lipidic endarteritis.

The etiology is probably autoimmune. Riedel's thyroiditis is frequently associated with elevated antithyroid antibody titers, and can be associated with concomitant Hashimoto's or Graves' disease. The mechanism of fibroblastic proliferation may relate to fibroblastic stimulation by anti-TSH receptor antibodies, as is seen with Graves' pretibial edema and ophthalmopathy.

The clinical presentation can be indistinguishable from malignancy. Radiographic imaging (magnetic resonance imaging or computed tomography) will reveal a thyroid mass with obliterated fat planes. Fine-needle aspiration is not likely to be diagnostic because the fibroblastic proliferation makes aspiration difficult. Core or open biopsies should establish the correct diagnosis. Histologically, a dense fibrosclerotic process and extensive chronic inflammation are seen. Fibroinflammatory vascular changes also may be seen.

Riedel's thyroiditis is usually a self-limiting condition. High-dose prednisone can produce dramatic improvement, and continued glucocorticoid administration is recommended to prevent progressive multifocal fibrosclerosis.

ECTOPIC THYROID

Cervical Lymph Node Inclusions versus Metastatic Disease

The issue of thyroid inclusions within cervical lymph node is controversial. Many pathologists consider the presence of any thyroid follicles in a lymph node to be evidence of metastasis, even in the absence of an established thyroid primary, because it is well known that papillary thyroid carcinoma is frequently announced by cervical metastasis. However, documented benign thyroid inclusions can, and do, occur within jugular lymph nodes. They represent entrapped remnants of thyroid tissue that have migrated during development. This gave rise to the term *lateral aberrant thyroid*. Benign inclusions are extremely rare, though. Thyroid tissue within cervical lymph nodes is almost always more likely to be due to metastatic disease rather than developmental inclusion.

The diagnosis of benign inclusions requires defined, strict criteria: (1) follicles should be located in only the lymph node capsule or subcapsular tissue; (2) there should be only a few follicles; (3) cytologically, they should be identical to normal thyrocytes, lacking the nuclear features of papillary thyroid carcinoma; and (4) the involved nodes should be from the jugular nodes, not paratracheal or supraclavicular. If more than one third of the lymph node is replaced by thyroid tissue or if any psammoma bodies, papillary structures, or nuclear features of papillary thyroid carcinoma are seen, metastatic papillary thyroid carcinoma is likely. Of course, any incidental findings of thyroid follicles within cervical lymph nodes must be investigated further.

Thyroid ultrasound with aspiration cytology or biopsy is an invaluable tool. If a thyroid nodule is found, lobectomy is required. If the workup is negative, and the above histologic criteria for benignity are met, no further surgery may be needed, although this is a rare situation. If the thyroid inclusions meet the criteria for malignancy, ipsilateral thyroid lobectomy is mandatory, whether or not a thyroid nodule is found. These cases almost always reveal occult microscopic papillary carcinomas. For this reason, the term *lateral aberrant thyroid* should be avoided in the context of neoplasia, because it does not remove the need to rule out primary carcinoma from the ipsilateral thyroid lobe. In rare instances the primary site may be in the opposite lobe.

Lingual Thyroid

Embryologically, the thyroid anlage originates from the foramen cecum, at the base of the tongue. This tissue descends as the thyroglossal duct, in the midline of the neck, reaching its normal position anterior to the trachea and larynx. Abnormalities in descent can occur anywhere along this route, and maldescent also may occur into the mediastinum. Lingual thyroid occurs when there is no, or only partial, descent of the thyroid from the foramen cecum. It is a rare clinical problem, but reported incidence varies from 1:10 to 1:100,000. Most cases occur in young women. Patients can present with dysphagia, dyspnea, and a hemorrhagic mass. Symptoms may increase with hormonal fluctuations (e.g., menstrual cycle) or may worsen with pregnancy. Patients are also frequently hypothyroid.

Grossly, the tissue has a spongy red appearance, and histologically, the follicles are similar to eutropic thyroid tissue but may be in a microfollicular arrangement. The follicles may appear to infiltrate into surrounding skeletal muscle, but they can be distinguished from the follicular variant of papillary thyroid carcinoma by the lack of diagnostic nuclear findings (see below).

Lingual thyroid may be diagnosed by biopsy and further confirmed by thyroid scan. If the diagnosis is established clinically, no surgery may be required; this may be the only thyroid tissue present in the patient, and the incidence of malignancy within lingual thyroid is extremely low. Symptomatic lingual thyroid may be managed by thyroid suppression, which will decrease the size of the lingual mass, and also supplants endogenous thyroid function.

Thyroglossal Duct Cyst

Thyroglossal duct cysts are dilated remnants of the path of the thyroid's descent. They are the most common cause of a midline mass of the upper neck. These cysts usually come to clinical attention in children or young adults, but they may suddenly appear in older patients. Patients present with a soft, cystic midline mass that moves with swallowing or with a thrusting of the tongue forward. Sometimes, pressure on the mass can result in discharge through communication with the foramen cecum, and patients will notice a bitter taste (dysgeusia) in their mouth. The cysts are almost always adherent to the hyoid bone, frequently penetrating it centrally and continuing upward to its lingual origins at the tongue base. They also can be present off the midline if the distal cystic portion is displaced laterally. The Sistrunk procedure, which includes resection of the central portion of the hyoid bone, is now standard practice and has virtually eliminated recurrences. Grossly, these cysts can average from 1 to 3 cm in diameter. The lining may have a shaggy appearance from repeated infections. Histologically, the cyst is epithelial lined with thyroid follicles in the cyst wall. Lymphoid infiltrate or fibrosis may obliterate the residual thyroid tissue. Approximately 1% of cases will contain a malignancy, usually papillary thyroid carcinoma.

Intratracheal Thyroid

Intratracheal thyroid rests represent one of the rarest forms of thyroid ectopia. A 2% incidence of incidental, subglottic, submucosal thyroid tissue has been observed in 250 laryngectomy specimens, with a 4:1 preponderance for left-sided trachea. These subglottic rests can cause dyspnea, or they may be incidental findings in the setting of thyroid carcinoma. The latter situation can lead to some confusion regarding tumor staging; a neoplasm may be erroneously upstaged if incidental intratracheal rests are interpreted as tumor extension. Because the trachea is posterior to the path of the descending thyroid, the mechanism of intratracheal ectopia is not immediately obvious. There are two prevailing, nonexclusive theories. The malformation theory states that thyroidal descent and formation are completed prior to tracheal cartilage formation, allowing for thyroid tissue to become entrapped and displaced by developing trachea. The purported left-sided propensity of intratracheal thyroid has been thought to relate to migratory differences between the left and right lateral ultimobranchial contribution to the thyroid anlage. The invasion theory states that thyroid tissue continues to migrate, albeit along an aberrant pathway, and becomes situated in the trachea as a result of direct, yet oncologically benign, invasion, as is seen quite frequently with mediastinal thyroid tissue.

Mediastinal Thyroid

Most cases of mediastinal (substernal) thyroid are inferior extensions of multinodular goiters and can be removed through the neck because their vascularization is derived cervically from the inferior thyroid artery. Truly aberrant mediastinal thyroid, which may be contained within the thymus or mediastinum, is rare, but cases often have a separate vascularization from mediastinal vessels or branches of the internal mammary artery, and therefore cannot be resected via cervical approach. Mediastinal goiter can compress the trachea or great vessels, causing respiratory failure or superior vena cava syndrome. Most patients are euthyroid, but uncommonly they can be hyperthyroid. The thyroid is usually markedly enlarged, multinodular, and histologically identical to typical multinodular goiters. In a small percentage of cases, a papillary carcinoma is present.

MALIGNANT THYROID NEOPLASIA

Papillary Carcinoma: Diagnosis, Pathobiology, and Variants

Papillary thyroid carcinoma is the most common form of thyroid malignancy, constituting 80% of all thyroid carcinomas. It occurs over a wide age range (5–95 years, median 39) with a female to male ratio of 3:1. There are some genetic associations for papillary thyroid carcinoma; it may be associated with familial adenomatous polyposis (Gardner's syndrome) and also Cowden's syndrome. Ionizing radiation, particularly during childhood for benign conditions such as thymic "enlargement," eustachian tube obstruction, and tinea capitus, is a predisposing factor. This can occur with accumulation of lower sublethal energy doses (e.g., 0.2 to 40 rads). In vitro cell line studies have demonstrated that high doses (10, 50, and 100 Gy) of x-rays produce preferential activation of the *RET/PTC1* oncogene (as compared with *RET/PTC2, RET/PTC3* induction) in a dose-dependent manner, confirming the initiating potential of ionizing radiation, as well as the significance of *RET/PTC1* rearrangement in the early steps of thyroid carcinogenesis. Although low doses of beam therapy in children and adolescents incite future thyroid cancer, which usually develops 1 to 2 decades later, high-dose beam therapy to the neck or chest, such as that for Hodgkin's disease, does not do so. There have been no reports of an increased incidence of thyroid cancer after high-dose beam therapy, although an increased risk of lung, salivary, and breast cancers, as well as sarcomas and other malignancies, has been noted. An increased risk for subsequent benign as well as malignant thyroid lesions also is seen after exposure to increased levels of radioactive iodine 131 fallout. This has been observed in survivors of the atomic bomb attacks on Hiroshima and Nagasaki, Japan, Marshall Islanders exposed to the Bikini atoll atomic bomb trail, and more recently in people exposed to fallout following the disaster at the Chernobyl nuclear power plant near Kiev, Ukraine. On the other hand, therapeutic doses of radioactive iodine 131 have an ablative effect on the thyroid and are not associated with increased risk of subsequent thyroid malignancy.

Patients with papillary thyroid carcinoma may present with discrete, thyroid masses, with or without adenopathy, or enlarged lymph nodes but no apparent thyroid mass. A frequent presentation in younger (5–19 years, mean 16 years) individuals is that of an unexplained enlarged jugular lymph node without an obvious site of primary disease. These pathological nodes characteristically represent metastasis from an occult, ipsilateral, localized sclerotic papillary thyroid carcinoma. Papillary carcinoma can have diverse appearances upon sectioning the gland: localized sclerotic nodules, diffusely sclerotic, infiltrative tumors, or encapsulated tumors mimicking follicular adenomata. Their interior can have a "furry," dull appearance, sometimes with yellow flecks, reflecting the relative lack of colloid and the presence of papillary projections (**Fig. 55–1**). The localized sclerotic form can be quite small (<1 cm), and nonpalpable (occult), coming to clinical attention because of the appearance of cervical metastases.

Figure 55–1 *Papillary thyroid carcinoma. This circumscribed tumor has a characteristic "furry" cut surface, indicating papillae formation.*

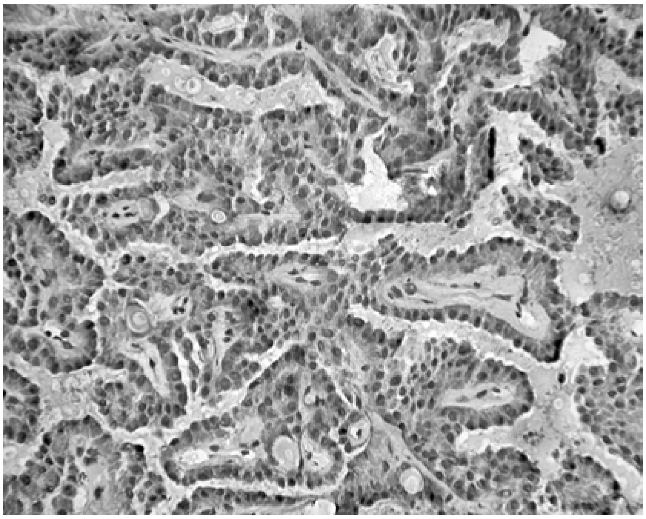


Figure 55–2 *Papillary thyroid carcinoma. This frozen section is diagnostic for papillary thyroid carcinoma, as abundant "true" papillae are formed with fibrovascular cores. Note, however, that the tumor nuclei are condensed; also, no "Orphan Annie eye" cleared nuclei are observed in frozen sections of this tumor.*

Papillary thyroid carcinoma has characteristic microscopic nuclear features. The nuclei are large and ovoid, as compared with normal thyrocytes, with a crowded overlapping appearance. Nuclear holes, grooves, and cleared nuclei are characteristic, but not specific or reliable, for the diagnosis. These findings are difficult to demonstrate at frozen section. The pathologist can rely only on the tumor architecture to make the diagnosis at that time (**Fig. 55–2**). Papillary thyroid carcinoma forms papillary structures, with fibrovascular cores, psammoma bodies, and follicular structures. If these follicular structures predominate histologically, the tumor is referred to as a follicular variant of papillary thyroid carcinoma. Although true tumor papillae can be discerned on frozen section analysis, the follicular variant of papillary carcinoma may be impossible to diagnose on frozen section, for not only are the diagnostic nuclear features lacking, the characteristic papillary architecture is now also lacking (**Fig. 55–3**). Previous aspiration cytology results can be helpful if papillary thyroid carcinoma had been definitively diagnosed. Thus most false-negative frozen section diagnoses for papillary thyroid carcinoma are, in fact, follicular variants of papillary thyroid carcinoma. Prior to the recognition of the follicular variant of papillary thyroid carcinoma, many of these tumors were misdiagnosed as follicular carcinomas.

Certain variants of papillary thyroid carcinoma deserve mention because they are associated with a poorer prognosis. A carcinoma that presents as a

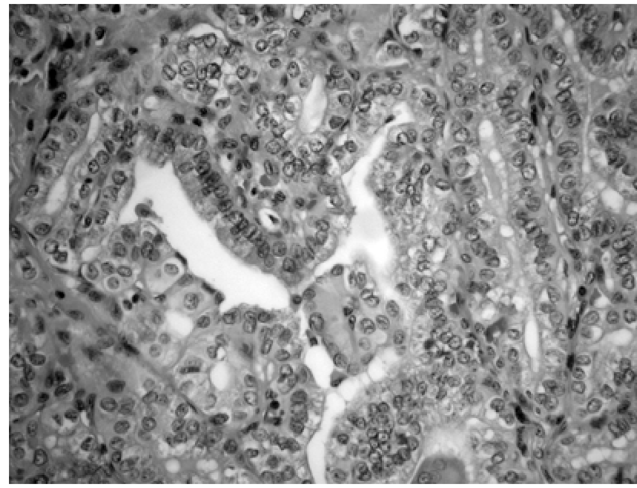


Figure 55–3 *Papillary thyroid carcinoma, follicular variant. This neoplasm reveals follicular structures, but with nuclear features diagnostic for follicular variant papillary thyroid carcinoma.*

diffusely sclerotic tumor will invariably present at a higher stage (T_3 or T_4), and is more aggressive. The tall-cell variant of papillary thyroid carcinoma represents ~10% of papillary carcinomas; it is defined as a tumor in which more than 30% of the cells are twice as tall as they are wide. It is frequently associated with high-grade features such as tumor necrosis, vascular invasion, solid areas, and increased mitotic figures. Several investigators found the tall-cell variant to have a more aggressive behavior than ordinary well-differentiated papillary carcinoma. Columnar cell carcinoma is a rare variant, defined by tall tumor cells with stratified nuclei and subnuclear vacuoles, and is also associated with more aggressive behavior. Papillary thyroid carcinoma may be an incidental synchronous finding of lobectomy performed for another nodule, such as an adenoma. These incidental occult tumors are usually small (<1 cm). If confined to the thyroid parenchyma, lobectomy is curative, and a complete thyroidectomy is not necessary.

The overall survival statistics for patients with papillary thyroid carcinoma are excellent; the Surveillance, Epidemiology, and End Results (SEER) program data reveal that the overall 10-year survival is 98%. This deceptively rosy survival rate reflects a preponderance of "typical" papillary thyroid carcinomas, which are classified as "low-risk" cases. Low-risk patients are young (<45 years), with small tumors (less than 4 cm in diameter) and no distant metastasis. These tumors also lack high-grade histological features (diffuse sclerotic variant, tall-cell variant) and do not extend into perithyroid soft tissue. The acronym AGES refers to variables impacting on prognosis (*a*ge >45, male *g*ender or tumor *g*rade, *e*xtrathyroid *e*xtension, and patient stage). Histological features that define high-grade papillary thyroid carcinoma include extrathyroid extension, diffuse sclerotic type, tall-cell variant, and columnar cell variant.

Twenty-year survival for intermediate risk groups is 83%, and high-risk is 43%. Distant metastases (e.g., lung) are associated with poor prognosis. The issue of regional, cervical metastases of papillary carcinoma is intriguing. Microscopic metastatic papillary carcinoma is common within cervical lymph nodes of patients with papillary carcinoma and has been documented in up to 80% of dissections. Unlike other malignancies, in which survival dramatically drops by 50% with the presence of metastatic lymph nodes, positive nodal disease conveys no survival impact for patients with papillary thyroid carcinoma. This unique quirk of papillary carcinoma was confirmed in a report on ~12,000 National Cancer Institute study patients, which demonstrated that cervical lymph node metastasis was not an independent prognosticator, and did not affect overall survival for patients with well-differentiated carcinoma. Performing jugular and central compartment lymph node dissection in the absence of clinical adenopathy has not been demonstrated to improve patient survival. The concept of "stagnant metastatic deposits" has been proposed because in most patients papillary thyroid carcinoma almost never leaves the region of the neck. When papillary thyroid carcinoma does metastasize to distant sites, it usually involves the lungs or brain.

Follicular Carcinoma

Follicular cell carcinomas represent 17% of thyroid malignancies. There is a wide age range (8–98 years), but the mean patient age at diagnosis (48 years) is a decade older than that of papillary thyroid carcinoma (39 years). A female predominance (73%) is present. These carcinomas can be subdivided into two subgroups: minimally invasive, well-differentiated encapsulated follicular carcinoma and invasive, well- to moderately differentiated follicular carcinoma. The presentation of a minimally invasive follicular carcinoma is identical to that of a follicular adenoma: patients present with a dominant nodule that is "cold" on a radioiodine scan. Fine-needle aspiration is of no help in distinguishing between these two entities; the diagnosis must be made on architectural, not cytological, features.

Patients with invasive follicular carcinoma can present with an irregular firm thyroid mass that may be fixed to the surrounding tissues. Occasionally, metastases may be the initial presentation, especially bony metastases. Here, preoperative cytology, in the clinical setting of an obvious clinical malignancy, can be helpful in distinguishing follicular carcinoma from papillary and anaplastic carcinomas.

Minimally Invasive Encapsulated Follicular Carcinoma

This is an encapsulated tumor, but the capsule is usually thicker than that of a routine adenoma. The interior of the tumor is usually mahogany colored but glistening, not "furry," as in papillary carcinoma. Histologically, the thyroid follicles are usually small, but bland and indistinguishable from those benign follicular adenomas. The diagnosis is established on permanent section by finding either complete penetration of the neoplasm through the capsule or subcapsular vascular/lymphatic invasion. If tumor extends into the thyroid parenchyma or necrosis is seen, then the tumor may be beyond "minimally invasive," and classified as invasive. Additional deeper sections from the paraffin block may be required for the diagnosis; the malignant thyrocytes may appear identical to benign thyrocytes, and the pathologist cannot rely on the usual cytological indicators of malignancy.

In a large ($n = 95$) series of minimally invasive encapsulated follicular carcinomas reported from the Armed Forces Institute of Pathology, the female to male ratio was 2.4:1, and patient ages ranged from 20 to 95 years (average 42.0 years). The mean mass diameter was 2.8 cm. All of these patients were treated by surgical excision, and adjuvant radioactive iodine therapy was administered to 24 patients. The prognosis of minimally invasive encapsulated follicular carcinoma is excellent. The recurrence rate was 5% (mean 18.1 years), with no disease-related mortality. The remaining patients were disease free after a significant follow-up period (mean 16.5 years), which is significantly better than is seen for invasive follicular carcinoma (see next section).

Invasive Follicular Carcinoma

Thyroid follicular carcinoma is the second most common thyroid malignancy, after papillary carcinoma. Hürthle cell carcinoma, insular carcinoma, and trabecular-insular carcinoma are histologic variants of invasive thyroid follicular carcinoma; all of these carcinomas have the same biological potential stage for stage. The true incidence of follicular carcinoma is difficult to determine from the older literature because many cases of follicular variant papillary thyroid carcinoma (see earlier) had been classified as follicular carcinomas. One retrospective study documented that the ratio of papillary to follicular carcinomas diagnosed varied from 0.60 in 1974–1976 to 6.88 in 1992–1994. This casts doubt on the survival statistics of follicular carcinoma in the older literature because the prognosis of the follicular variant of papillary carcinoma is identical to that of papillary thyroid carcinoma, which is usually better than that of invasive follicular carcinoma. Marked geographical variations in the relative ratios between follicular and papillary thyroid carcinoma have been observed, probably relating to dietary iodine deficiencies.

On clinical examination, invasive follicular carcinoma presents as a firm thyroid mass. Fixation of the surrounding soft tissues raises the clinical suspicion of malignancy. A small percentage of patients (up to 11%), may present with metastatic disease (usually within bone, lungs, and soft tissues). Histologically, this form of follicular carcinoma can be either well or moderately differentiated, but it invariably demonstrates frank invasion into the thyroid parenchyma, as well as vascular invasion. Invasion of extrathyroid soft tissue may be seen and is a poor prognostic feature. The tumor cells of well-differentiated follicular carcinoma appear similar to normal thyrocytes, but they may be larger, with greater atypia. They can be distinguished from papillary thyroid carcinoma tumor cells because their nuclei are round and nonoverlapping. The tendency toward vascular invasion also suggests the diagnosis of follicular carcinoma. Histologic variants of follicular carcinoma include trabecular-insular, Hürthle cell, clear cell, and mixed medullary-follicular carcinoma. These variants can be more solid and moderately to poorly differentiated.

The issue of Hürthle cell carcinomas has been shrouded in some mystery due to some early publications by surgeons of "Hürthle cell adenomas" that metastasized. Subsequent pathological studies have confirmed that the diagnostic criteria necessary to distinguish Hürthle cell adenoma from carcinoma are the same as in other follicular tumors; namely, capsular invasion and/or subcapsular vascular invasion (**Fig. 55–4**). As mentioned, large Hürthle cell tumors (>4 cm) are more likely to be malignant than small tumors (<2 cm).

Hürthle cell carcinomas do appear inherently different in some ways from follicular cancers; they rarely take up radioactive iodine but do make thyroglobulin, and they more commonly spread to lymph nodes. A review from Memorial Sloan Kettering Cancer Center of over 1000 thyroid malignancies demonstrated that those tumors diagnosed as Hürthle cell carcinoma have a greater likelihood of developing distant metastases than follicular carcinomas. However, regression analysis stratifying for tumor stage was lacking in this article, so further clinicopathologic study is necessary before this issue can be confirmed. The data published from the SEER program (1973–1991) do not examine this issue.

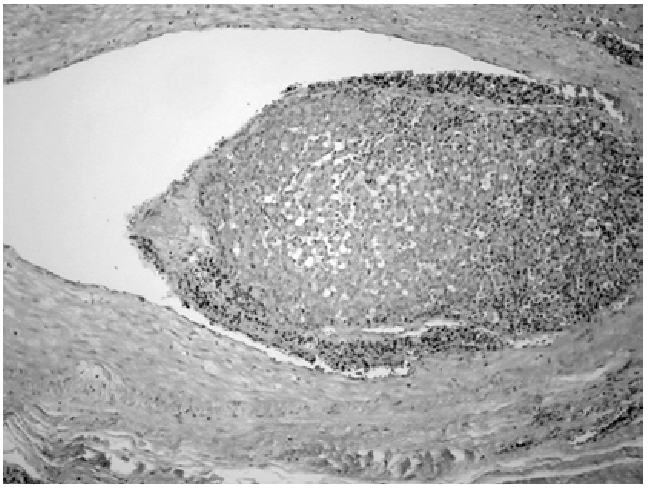


Figure 55–4 *Follicular carcinoma. Tumor emboli are present within capsular vascular spaces.*

However, the National Cancer Institute data do actually stratify follicular thyroid carcinoma and Hürthle cell carcinoma by AJCC stage. This report finds that, whereas Hürthle cell carcinomas are less likely to be encountered as stage I malignancies, stage for stage, patient survival appears similar for follicular carcinoma and Hürthle cell carcinoma.

Immunohistochemistry may be helpful to distinguish some solid follicular carcinoma variants, such as trabecular-insular variant (which invariably express thyroglobulin) from medullary carcinoma (which will express calcitonin, and possibly other neuroendocrine markers, but not thyroglobulin).

The prognosis of invasive follicular carcinoma is much poorer than that of minimally invasive follicular carcinoma or papillary thyroid carcinoma. The reported mortality rates vary from 16 to 60%. Several studies have identified poor prognosticators. Older patients (>45 years), presence of distant metastases, thyroid capsular invasion, and tumor size greater than 2 cm are features invariably associated with a poorer prognosis.

Medullary Carcinoma

Medullary thyroid carcinoma represents ~3% of all thyroid malignancies. Most cases (80%) occur sporadically, whereas 20% of these tumors are familial and may be associated with other endocrine pathologies such as pheochromocytoma and hyperparathyroidism [multiple endocrine neoplasia type 2A (MEN2A), MEN2B, and isolated familial medullary thyroid carcinoma]. The average age of presentation for sporadic tumors is 52.3 years, whereas the average age for index patients with hereditary tumors is 29 years and for all screened patients is 23 years. Within this last group, the median age of diagnosis for hereditary tumors varies: 36 years for hereditary medullary carcinoma, 27.4 years for MEN2A, and 17.2 years for MEN2B. Patients with sporadic disease usually present with palpable thyroid masses. The index patient with familial disease (the first individual in a cohort to be identified) can present with a thyroid mass but also may be hypertensive secondary to pheochromocytoma or hypercalcemic (with its attendant symptoms) due to hyperparathyroidism. Ultrasound and fine-needle aspiration should then be performed for possible preoperative diagnosis of medullary carcinoma. Any suspicion for the diagnosis of medullary thyroid carcinoma should be relayed to the cytologist so that provision can be made to perform immunohistochemistry for calcitonin. Whereas immunohistochemistry can be performed retrospectively from the formalin-fixed, paraffin-embedded tissue blocks, cytology specimens require advance preparation of either additional unstained slides or a "spun down" cytology block for immunohistochemistry.

The pentagastrin stimulation test can then confirm elevated calcitonin, hence the diagnosis. The provocative pentagastrin stimulation test is useful in distinguishing medullary carcinoma from false-positive, mild elevations in serum calcitonin. The preoperative diagnosis of medullary carcinoma should initiate investigations to distinguish sporadic tumors from index familial cases. The latter patients then require screening for other potential endocrine tumors, as well as intensive familial screening. The identification of germline mutations in the proto-oncogene *RET* identifies these familial cases; this blood testing is especially important for all patients younger than 45 years. Within identified families, asymptomatic members can be screened and identified by elevated serum calcitonin. Medullary carcinoma is associated with an overall mild female predominance, which is more pronounced in sporadic cases (female to male ratio 1.84:1) than in hereditary cases (female to male ratio 1.18:1).

Medullary carcinomas arise from the neuroendocrine parafollicular C cells (calcitonin-secreting C cells). These neural crest–derived cells are more concentrated within the lateral and superior aspects of the thyroid lobes. Not surprisingly, these tumors, especially the familial ones, have a predilection for the superolateral thyroid lobes. Grossly, they are usually circumscribed, with a smooth-cut surface, tan-white in color. Familial tumors may be smaller and multiple as compared with sporadic tumors, with adjacent C-cell hyperplasia (see later discussion). Histologically, medulllary carcinoma is composed of epithelioid or spindled neuroendocrine cells. Their nuclei have the fine, stippled chromatin characteristic of neuroendocrine tumors. Calcitonin deposition can be seen by light microscopy as amyloid and confirmed by immunohistochemistry. However, some medullary carcinomas can mimic "oat cell" (small-cell) carcinomas, in that they appear as amyloid-poor, malignant, small, blue round cells. The diagnostic immunohistochemical profile for medullary carcinoma is expression of calcitonin and carcinoembryonic antigen, but no thyroglobin expression. C-cell hyperplasia represents a histologically benign proliferation of C cells and is the earliest thyroid manifestation of this endocrinopathy. It appears as nests of calcitonin-positive epithelial cells adjacent to follicles, the size and number of which are increased as compared with normal controls.

Total thyroidectomy and central neck dissection are the minimum therapy for patients with medullary carcinoma. In over 90% of cases the disease is multicentric, and in contrast to well-differentiated thyroid cancer, positive nodes do have an adverse effect on survival. Prophylactic total thyroidectomy is recommended for all probands within identified families; it should be performed as soon as possible. Surgery is warranted as early as 6 years of age, and these thyroids usually reveal only C-cell hyperplasia. The prognosis of medullary thyroid carcinoma correlates with stage at presentation. Tumors detected by screening are identified at an earlier stage and are cured by total thyroidectomy and paratracheal lymph node dissection. Conversely, the initial presence of systemic symptoms (diarrhea, bone pain, or flushing) is usually associated with widely metastatic, sporadic disease, and a significant (i.e., 33.3%) 5-year mortality rate. Disease-free survival at 10 years is around 50% for sporadic medullary thyroid carcinoma.

Anaplastic Carcinoma

Anaplastic thyroid carcinoma comprises 5 to 15% of primary thyroid malignancies. It represents one of the most aggressive solid tumors in humans, usually presenting in the sixth and seventh decades of life, with a small female preponderance. The classic presentation is that of a rapidly growing neck mass causing hoarseness and airway obstruction. The acute symptoms frequently appear a short time before patients seek attention. More than half of patients will have a history of long-standing goiter. Leukocytosis and fever commonly accompany presentation. Distant metastases, usually to the lung, are common at presentation. Ultrasound and fine-needle aspiration cytology usually can establish the diagnosis preoperatively. If the aspiration is inconclusive, an open biopsy is required. Histologically, anaplastic thyroid

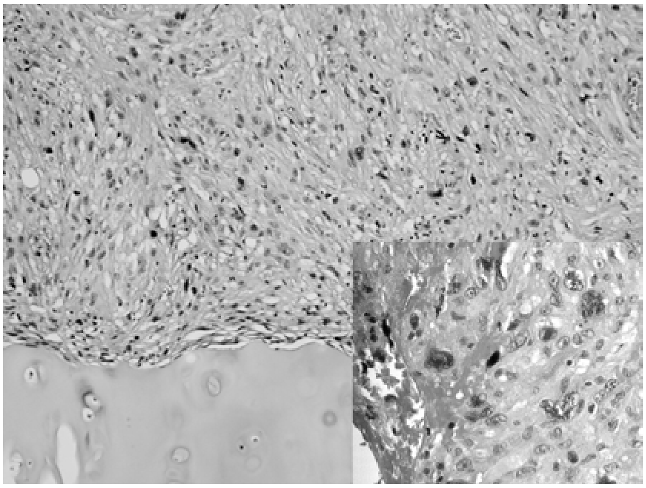


Figure 55–5 *Anaplastic carcinoma. This highly malignant neoplasm can be seen invading laryngeal cricoid cartilage Inset: Anaplastic carcinoma is characterized by extreme nuclear pleomorphism and the presence of spindled and multinucleated tumor giant cells.*

carcinoma is composed of huge, bizarre, anaplastic spindled cells and multinucleated tumor giant cells. The tumor cells are the most pleomorphic of any head and neck malignancies. The spindled cells resemble sarcoma (**Fig. 55–5**). Epithelioid tumor cells with squamous metaplasia can be present, which may be significant if a clinical differential diagnosis involves upper aerodigestive tract carcinomas. The majority of anaplastic thyroid carcinomas are found to contain foci of well-differentiated thyroid carcinoma, when adequately sectioned. Immunohistochemistry for thyroglobulin is usually negative in anaplastic carcinomas; thyroid transcription factor (TTF) is also usually negative, or it may be focally positive. If the anaplastic carcinoma is amenable to surgery, then it should be resected. Aggressive multimodality treatment radiation and chemotherapy should be offered in any case; however, survival rates remain low, and the median survival for most series is under 6 months.

OTHER THYROID TUMORS

Squamous Carcinoma: Primary versus Secondary

Primary thyroid squamous carcinoma is rare, accounting for 0.7 to 3.4% of all thyroid malignancies. Approximately 130 cases have been reported in the English literature. It arises in either of two settings: either as a primary tumor or as a component of anaplastic or tall-cell papillary thyroid carcinoma. There is a female predisposition, with a female to male ratio of ~2:1. Most patients are in the fifth to sixth decades of life, similar to that of anaplastic thyroid carcinomas. Presentation is dramatic; patients have rapidly enlarging neck swelling, shortness of breath, and hoarseness. The tumors have been described as firm, necrotic, and extensively infiltrating the thyroid and perithyroidal soft tissue; they may extend to the larynx, trachea, and mediastinum. Positive cervical lymph nodes are common at presentation, and some patients also may have pulmonary metastases upon presentation.

Histologically, primary thyroid squamous carcinoma is composed of islands of malignant squamous cells with varying degrees of differentiation. Areas of well-differentiated or tall-cell papillary carcinoma may be present but not as the dominant histological component; it should constitute less than 30% of the carcinoma. A spindle cell component can also be present. The main differential diagnosis includes direct extension of a carcinoma from the larynx or esophagus, or direct extension from a metastatic cervical lymph node from an upper aerodigestive tract primary. Clinical correlation, as well as finding papillary thyroid differentiation, or immunohistochemical expression of thyroid transcription factor or thyroglobulin, can establish a carcinoma as being of thyroid origin.

Squamous cell carcinoma of the thyroid gland is a highly lethal tumor. Most patients die within a year despite combination therapy including surgery, radiation, and chemotherapy. The cause of death is often related to local recurrence or distant metastases (lungs, liver, heart, kidneys). The prognosis is similar to or even worse than anaplastic thyroid carcinoma.

One variant of thyroid squamous carcinoma that requires mention is CASTLE (*c*arcinoma *s*howing *t*hymus-*l*ike *d*ifferentiation). CASTLE tumors of the thyroid are rare and indolent with a favorable prognosis. Histologically, they are lobulated and expansive, with fibrous septa and indistinct cellular borders, and may have foci of squamous differentiation.

Metastatic Disease to the Thyroid

Metastatic malignancies to the thyroid are relatively rare. In many cases, the metastasis may be the initial manifestation of disease. The most common sources of primary disease are lung, gastrointestinal tract, and melanoma. Carcinomas from prostate, larynx, kidney, and breast also may metastasize to the thyroid.

FROZEN SECTION ANALYSIS: STRENGTHS AND LIMITATIONS

Intraoperative tissue analysis involves sampling and rapidly freezing tissue that can be cryostat sectioned. The tissue sections are immediately stained and examined microscopically. The tissue fixation step, which preserves morphology, is technically not feasible during frozen section.

Thus the biggest limitation of frozen section analysis is suboptimal tissue morphology. Frozen section analysis will allow for the diagnosis of papillary thyroid carcinoma with typical papillary architecture, invasive follicular carcinoma, Hashimoto's thyroiditis, anaplastic carcinoma, and multinodular goiter. There are three common areas where frozen section analysis meets its limitations: (1) distinguishing follicular variant of papillary thyroid carcinoma from a follicular adenoma, (2) definitively identifying minimally invasive follicular carcinoma, and (3) distinguishing the papillary hyperplasia of Graves' disease from papillary carcinoma. Because these first two issues are generally encountered while examining follicular adenomas, one can imagine just how often this differential diagnosis does arise.

When one reexamines false-positive or false-negative frozen section diagnoses regarding papillary thyroid carcinoma, these errors usually involve misdiagnosing follicular variant of papillary thyroid carcinoma. As mentioned, when papillary architecture is lacking, one can only base the diagnosis of follicular variant papillary thyroid carcinoma on nuclear morphology, which is compromised at frozen section analysis. The lack of formalin fixation obscures the nuclear holes, grooves, and cleared nuclei characteristic for the diagnosis. Hence a surgeon wishing to perform a total thyroidectomy after a definitive diagnosis of papillary thyroid carcinoma must be patient and await the permanent section diagnosis 48 hours later.

Minimally invasive follicular carcinoma, as mentioned, is an uncommon, low-grade malignancy. The diagnosis is based on observing either capsular invasion or subcapsular vascular invasion in a solitary follicular nodule. Grossly, these lesions resemble follicular adenomas. However, it has been noted that the fibrous capsules of these tumors can be thicker than those of conventional adenomas. Many histologic samples may be necessary before a diagnosis of minimally invasive follicular carcinoma is made. However, sampling limitations at the time of frozen section usually preclude this diagnosis. Importantly, these minimally invasive follicular carcinomas are cytologically identical to follicular adenomas. Theoretically, any apparent follicular adenoma may yield diagnostic evidence of invasion after adequate sampling. For this reason, most pathologists will not definitely diagnose a benign follicular adenoma on frozen section. The diagnosis of "histologically benign follicular neoplasm" is used to imply that the lesion is consistent with follicular adenoma, but that definitive diagnosis awaits the permanent sections.

Finally, Graves' disease is usually treated with thyroid suppression. Surgery may become necessary in the course of this illness if a thyroid nodule develops. Alternatively, a follicular adenoma may be hyperfunctioning (e.g., "hot"). The appearance of a hyperfunctioning adenoma is identical to Graves' disease, with papillary hyperplasia. Under these conditions, the pathologist may histologically encounter the papillary hyperplasia of Graves' disease and need to distinguish it from papillary carcinoma. Again, the lack of nuclear details inherent to frozen section can lead the pathologist to defer definitive diagnosis until permanent section.

Suggested Readings

Brandwein M, Som PM, Urken M. Benign intratracheal thyroid: a possible cause for preoperative overstaging. Arch Otolaryngol Head Neck Surg 1998;124:1266–1269

Derringer GA, Thompson LD, Frommelt RA, et al. Malignant lymphoma of the thyroid gland: a clinicopathologic study of 108 cases. Am J Surg Pathol 2000;24:623–639

Evans HL. Follicular neoplasms of the thyroid: a study of 44 cases followed for a minimum of 10 years, with emphasis on differential diagnosis. Cancer 1984;54:535–540

Farahati J, Demidchik EP, Biko J, Reiners C. Inverse association between age at the time of radiation exposure and extent of disease in cases of radiation-induced childhood thyroid carcinoma in Belarus. Cancer 2000;88:1470–1476

Gilliland FD, Hunt WC, Morris DM, Key CR. Prognostic factors for thyroid carcinoma: a population-based study of 15,698 cases from the Surveillance, Epidemiology and End Results (SEER) program, 1973–1991. Cancer 1997;79:564–573

Giuffrida D, Gharib H. Anaplastic thyroid carcinoma: current diagnosis and treatment. Ann Oncol 2000;11:1083–1089

Hundahl SA, Fleming ID, Fremgen AM, Menck HR. A National Cancer Data Base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985–1995. Cancer 1998;83:2638–2648

Johnson T, Lloyd RV, Thompson NW, et al. Prognostic implications of the tall cell variant of papillary thyroid carcinoma. Am J Surg Pathol 1988;12:22–27

Lam KY, Lo CY, Liu MC. Primary squamous cell carcinoma of the thyroid gland: an entity with aggressive clinical behaviour and distinctive cytokeratin expression profiles. Histopathology 2001;39:279–286

Leteurtre E, Leroy X, Pattou F, et al. Why do frozen sections have limited value in encapsulated or minimally invasive follicular carcinoma of the thyroid? Am J Clin Pathol 2001;115:370–374

LiVolsi VA. Hyalinizing trabecular tumor of the thyroid: adenoma, carcinoma, or neoplasm of uncertain malignant potential? Am J Surg Pathol 2000;24:1683–1684

Meyer JS, Steinberg LS. Microscopically benign thyroid follicles in cervical lymph nodes: serial section study of lymph node inclusions and entire thyroid gland in 5 cases. Cancer 1969; 24:302–311

Modigliani E, Cohen R, Campos JM, et al. Prognostic factors for survival and for biochemical cure in medullary thyroid carcinoma: results in 899 patients: the GETC Study Group (Groupe d'étude des tumeurs a calcitonine). Clin Endocrinol (Oxf) 1998;48:265–273

Ng WK, Collins RJ, Shek WH, Ng IO. Cytologic diagnosis of "CASTLE" of thyroid gland: report of a case with histologic correlation. Diagn Cytopathol 1996;15:224–227

Shaha AR, Shah JP, Loree TR. Risk group stratification and prognostic factors in papillary carcinoma of thyroid. Ann Surg Oncol 1996;3:534–538

Thompson LD, Wieneke JA, Paal E, et al. A clinicopathologic study of minimally invasive follicular carcinoma of the thyroid gland with a review of the English literature. Cancer 2001;91:505–524

Wenig BM, Thompson LD, Adair CF, et al. Thyroid papillary carcinoma of columnar cell type: a clinicopathologic study of 16 cases. Cancer 1998;82:740–753

Wirtschafter A, Schmidt R, Rosen D, et al. Expression of the *RET/PTC* fusion gene as a marker for papillary carcinoma in Hashimoto's thyroiditis. Laryngoscope 1999;109:1011–1015

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. Medullary carcinoma
   A. Is the most common thyroid malignancy diagnosed in the first 2 decades of life
   B. Is most commonly associated with a somatic mutations of the *RET* proto-oncogene
   C. Requires the finding of amyloid for the diagnosis
   D. All of the above
   E. None of the above
2. Hürthle cell carcinomas
   A. Are usually inherently more aggressive than thyroid follicular carcinoma
   B. Are less likely to metastasize to bone than tall-cell variant of papillary thyroid carcinoma
   C. Are usually less than 2 cm when diagnosed
   D. Can be distinguished from Hürthle cell adenomas by fine-needle aspiration
3. Which of the following statements is true regarding papillary thyroid carcinoma (PTC)?
   A. The follicular variant of PTC is inherently more aggressive than the usual PTC.
   B. PTC is more aggressive in patients under 40 years of age.
   C. PTC is associated with therapeutic radioactive iodine exposure.
   D. Its incidence is increased in patients with Hashimoto's thyroiditis.

This page intentionally left blank

Part VI
FACIAL PLASTICS AND MISCELLANEOUS

Chapter 56

IMAGING OF THE NECK

ADAM SILVERS

Imaging the neck requires close cooperation between the clinician and the radiologist. In most cases, the superficial site and histology of the lesion are known prior to imaging, and the preoperative radiographic study is directed toward accurate mapping of the primary lesion, especially the presence of any deep extension, as well as the evaluation of any possible nodal metastasis. Thus the goal of such imaging is to provide the clinician with the most accurate tumor assessment so that the best treatment plan can be initiated. The imaging emphasis is placed on those areas that are difficult to assess clinically so that the combined radiographic and clinical evaluation is the best and most informed appraisal possible. Presently there are many imaging choices available, and in the current economic climate, consideration must be given not only to the best imaging modality but also to the most cost-effective choice. Thus the clinician–radiologist dialogue has at least two important aims: just as the radiologist depends on the referring clinician for insights into the physical findings, the imager should provide the clinician with recommendations as to appropriate choices for imaging.

IMAGING TECHNIQUES

As technology has improved, the older radiographic techniques are giving way to the newer imaging modalities. Thus plain films are no longer the primary imaging approach, and computed tomography (CT) and magnetic resonance imaging (MRI) have emerged at the forefront of these modalities. The techniques used include plain films, ultrasound, CT, and MRI.

PLAIN FILMS

Although routine plain films of the skull and mandible are easy to obtain from both a logistic and an economic standpoint, they no longer play a major role in the imaging of the neck. This primarily reflects their failure to assess soft tissue disease and, to a lesser degree, their ability to provide only a gross evaluation of the bone. However, plain film dental studies still remain the best method of imaging periodontal and periapical disease. Because the presence of such disease usually leads to complications if the patient receives radiation therapy, such affected teeth must be identified (and usually extracted) prior to treatment.

Dental studies consist of intraoral radiographs, with the film being placed within the oral cavity against the teeth. The three types of intraoral films are the periapical, the bitewing, and the occlusal, with the occlusal films providing the most information regarding the adjacent mandible. Occlusal films are obtained by placing the film between the surface of the teeth parallel to the occlusal plane, then using a steep angulation of the x-ray tube. These views are helpful not only in evaluating mandibular cortical erosion but also in evaluating pathological fractures of the mandible.

Panoramic radiography (Panorex, Imaging Sciences International, Hartfield, PA) has the advantage of visualizing the entire mandible on a single film; however, the resolution of a Panorex film is less than that of a routine dental film, and there is image distortion both in the midline and far laterally near the condyles. With recent improvements in CT image resolution and its ability to reformat images in multiple planes [i.e., Dentascan-type software programs (GE Healthcare Technologies, Waukesha WI)], the reliance on these specialized plain film techniques also has decreased. Although the inferior alveolar canal can be assessed on Panorex, other dental plain films, and Dentascan images, it is probably best evaluated on the Dentascan views where the canal can be seen in cross section.

Ultrasound

The basis of ultrasound imaging is the computerized conversion of reflected sound waves into diagnostic images. The advantages of ultrasound include its noninvasive nature, its lack of ionizing radiation, and the fact that it is in general less expensive than CT or MRI. However, the disadvantages of ultrasound include the limited penetration of its sound waves and imaging distortion due to disruption of the sound waves by both air and bone interfaces. Other disadvantages are that it is highly dependent on the skill of the user and the observation that most clinicians are uncomfortable with the ultrasound anatomy display. In the United States, ultrasound plays almost no role in the evaluation of the patient with head and neck cancer. However, recent reports from Europe have suggested that ultrasound examinations utilizing intraoral high-frequency fingertip-sized transducers may be helpful in evaluating small superficial oral carcinomas. Although such ultrasound may play a role in the mapping of small oral cancers, it has limitations in the evaluation of other upper aerodigestive tract disease, and it is not as reliable as CT and MRI for evaluating possible deeply situated metastatic adenopathy. Thus, in the United States, ultrasound remains a secondary study for the imaging of the neck, except for specific evaluation of the thyroid and parathyroid.

Computed Tomography

Computed tomography has been the dominant technique for imaging of the neck for the past 2 decades. The CT images are the result of using a highly focused x-ray beam that rotates around the patient. As the x-rays traverse the patient at multiple angles, computerized detectors analyze the degree that the tissues attenuate the beam. This information is then computer processed to provide a cross-sectional image of the patient. The data can be manipulated by choosing window and center settings to best evaluate bone (wide "bone" windows) or soft tissues (narrower windows), while the center or level setting, in effect, adjusts the picture brightness. In addition, specific filters (bone algorithm) can be used to achieve high-resolution bone images, and the combination of a bone algorithm processing filter with image display at "bone" windows provides the maximum CT information regarding osseous structures.

Software packages are presently available for most CT scanners that allow reformatting of the images in multiple planes. The General Electric Dentascan software package (or similar software programs from other vendors) provides both multiple panoramic images of the mandible and maxillary alveolus as well as multiple cross-sectional images of the jaws. The panoramic images are almost distortion free, are tomographic images through the bone, and are an excellent alternative to the conventional Panorex film. Although these dental software programs were mainly developed to aid in the planning of dental implants, these images also provide excellent information regarding the condition of the mandible in the patient with oral cancer. In particular, the multiple cross-sectional images (each perpendicular to the axial axis of the mandible) allow evaluation of the integrity of the mandibular lingual cortex, upper alveolus, and inferior alveolar canal. Software is also available that provides three-dimensional (3-D) image reconstruction from a series of axial or coronal CT scans. 3-D views can be particularly helpful after complex reconstruction of bony architecture. As with the Dentascan-type reformatting, the key to producing good reconstructions is the ability to obtain thin-section CT scans without patient movement. The computer from these initial axial or coronal studies generates the reformatted images, and patient movement causes misalignment artifacts in the final reconstruction; furthermore, the use of too thick CT sections causes poor resolution in the reconstructed images. Thus, when a

Dentascan or a 3-D reconstruction is needed, 1 to 3 mm contiguous CT scans should be obtained, usually in the axial projection because this position minimizes patient movement. Alternatively, today spiral CT can be used to obtain thin-section axial studies quickly. Although there is a slight loss of resolution when compared with a conventional axial scan, this is not evident in the reconstructed images.

The standard CT protocol used is a complete study of the neck performed as axial 3 mm contiguous sections from the level of the external auditory canal (skull base) to the level of the thoracic inlet. This approach allows evaluation of potential primary tumor sites and any cervical nodal metastasis. For this routine examination, the CT gantry is angled parallel to the infraorbital-meatal plane. Direct coronal 3 mm contiguous studies also can be obtained through the mouth with the patient in either a prone or a supine position. However, this requires that the patient can maintain this positioning for the duration of the examination, a feat not always possible due to pain, cervical spine disease, difficulty breathing, and so on. In addition, dental amalgams, if present, usually significantly degrade these coronal images.

Radiation dose to the lens of the eye for a routine axial and coronal CT is ~30 mGy, which is roughly 5 times the dose for standard x-rays of the sinuses. The thyroid gland and bone marrow within the mandible receive similar doses of ~30 mGy. The brain receives approximately half as much (14 mGy) during a routine CT.

If a Dentascan is performed, the CT gantry is angled perpendicular to the CT table, and the patient's head is positioned so that the CT beam is parallel to the occlusal surface of the teeth. An additional advantage of the Dentascan software package is that the mandibular and maxillary reconstructions can be limited to those images not degraded by the amalgams.

To improve potential contrast between tumor and normal tissues, it is best to perform the CT study using intravenous iodinated contrast material, which is administered using a power injector during the scan. Patients with a previous history of an allergic reaction to iodinated contrast material usually can tolerate the newer nonionic contrast materials without significant reaction. In more extreme cases, medicating the patient with steroids prior to the examination is usually effective and is recommended.

CT has several advantages over MRI (**Table 56–1**). It is a less expensive examination than MRI; claustrophobic patients who cannot tolerate MRI studies even with sedation (up to 10% of patients) usually can tolerate a CT scan; and metallic implants such as pacemakers, intracerebral aneurysm clips, and cochlear implants, which are contraindications for MRI, are not contraindications to CT. The size and weight limit restrictions for patients on MRI scanners are usually less restrictive on CT scanners. The CT study can be completed more rapidly than an MRI examination, CT scans are less sensitive to motion artifact than are MRI scans, and CT is better than MRI for evaluating cortical bone.

TABLE 56–1 CHOOSING IMAGING MODALITY

Advantages of CT	Advantages of MRI
Less expensive than MRI	No ionizing radiation
Quicker than MRI	Better tissue contrast
Less sensitive to motion	Can get multiple planes
Evaluates bone better	Less contrast reactions
Can fit larger patients	Less artifact from dental amalgam
Less claustrophobic	Better evaluation of bone marrow involvement
Easier to do on critically ill patients	Better tumor mapping than CT

CT, computed tomography; MRI, magnetic resonance imaging.

Ultrafast CT has been used on occasion to help evaluate the velopharyngeal region in patients with sleep apnea and in patients with velopharyngeal incompetence. However, the decreased image resolution (compared with conventional CT) has limited its usefulness in imaging the patient with cancer. The potential use of ultrafast CT was also considered to help evaluate patients with swallowing abnormalities. However, the best evaluation of swallowing problems remains the barium swallow study. This reflects the fact that the swallow occurs so rapidly that only a conventional barium study recorded on videotape (for slow-motion review) can capture any stepwise dysfunction that may be present.

MAGNETIC RESONANCE IMAGING

MRI scanners have been available in the United States for more than 20 years. During that time, the examination time has decreased significantly, image resolution has increased, and MRI has played an increasing role in the evaluation of head and neck cancer. The basis of MRI is very different from that of plain films and CT scanning in that there is no ionizing radiation. Rather, MRI images are obtained by measuring how rapidly hydrogen nuclei of different tissues return to their resting energy states after being excited by a strong magnetic field. This is done by first placing the patient in a strong magnetic field, which initially aligns the hydrogen nuclei in similar directions. Then a radiofrequency pulse is sent through the tissues, causing some of the hydrogen nuclei to resonate at a certain frequency, which can be detected as a radiofrequency wave by an external antenna or coil. By utilizing complex mathematical formulations, the MRI computer converts the differences in hydrogen nuclei relaxation times of the different tissues into cross-sectional images. The timing of the excitational pulse and data collection can be adjusted by using different pulse sequences (such as T1 and T2 weighted), which alter the relative relaxation times of one tissue to another. By obtaining multiple sequences, information regarding tissue differences is maximized. In general, T1-weighted sequences better demonstrate anatomy, and T2-weighted sequences better differentiate pathological tissue from normal tissue.

The addition of an inert paramagnetic intravenous contrast agent to an MRI examination can further help delineate pathological tissue from adjacent normal structures. The most often used contrast material is gadopentate dimeglumine (Magnevist, Berlex), commonly called gadolinium, which when absorbed by pathological tissue increases the signal intensity of the tissue as compared with skeletal muscle. This effect is seen mainly on T1-weighted sequences, and to a lesser degree on T2-weighted sequences. Because both fat and the enhanced tissue have bright signal on T1-weighted sequences, the use of fat suppression techniques (nulling the specific signal intensity of fat) on the postcontrast scans allows the enhancing tissue to be best identified.

The typical MRI examination of the neck usually consists of axial T1- and T2-weighted sequences, as well as a coronal T2-weighted study. Axial and coronal T1-weighted sequences with fat suppression are then obtained after intravenous contrast administration. Sagittal sequences are obtained as needed to better visualize the craniocaudad extent of disease. A routine examination can be obtained in approximately 30 to 40 minutes (vs 10 to 15 minutes for a CT study).

Advantages of MRI over CT include the fact that multiple planes can be imaged without changing patient positioning and that overall MRI offers greater contrast differences between pathological tissues and adjacent normal tissues. Although artifact from dental amalgams degrades MRI images, this effect is usually significantly less than that seen on CT scans. Furthermore, MRI does not utilize ionizing radiation, an important factor in the pediatric population.

In general, CT remains the initial preoperative examination for the patient with head and neck cancer, and MRI is usually obtained as a limited study to answer specific questions. MRI may better image tumor margins than was possible on CT and will better evaluate possible mandibular and skull base medullary (marrow) space invasion. However, with the continued decrease in MRI scan times, the increase in MRI resolution, and the gradual lowering of the examination cost, MRI is expected to play an increasing role in the preoperative evaluation of head and neck cancer. In the near future, real-time MRI scanning will be available, and so-called magnetic imaging fluoroscopy may allow the radiologist to better evaluate pharyngeal function. However, until this time, the barium swallow with videotape slow-motion review remains the best modality to evaluate deglutitional problems.

Peptide Receptor Scintigraphy

Nuclear medicine studies can be used as screening examinations, which may help focus cross-sectional imaging. One such technique is somatostatin analogue imaging. Somatostatin membrane receptors can be found in tumors of neuroendocrine origin such as paragangliomas and medullary thyroid carcinoma.

Indium-labeled octreotide is an analogue of somatostatin that has many of the features of somatostatin and binds to somatostatin receptor sites in the cell membrane. It can be injected into patients to help localize sites of disease.

In patients with paragangliomas, there is greater than 90% uptake rate of octreotide. In clinical studies octreotide scanning detected unsuspected additional paraganglioma sites in ~35% of patients with known paragangliomas. Somatostatin analogue imaging can also be helpful in patients with medullary thyroid cancer. There is approximately a 70% uptake rate with octreotide scanning; however, it may miss hepatic metastases.

In the future somatostatin analogues may be used for treatment of neuroendocrine tumors as targeted radionuclide therapy. Currently, the major use of these techniques is as a screening study that is followed by CT or MRI to better anatomically locate sites of suspected disease.

CHOICE OF IMAGING TECHNIQUE

The decision of which imaging modality to use as a first choice depends on the location of the tumor and the precise clinical/pathological information that is wanted. If a nasopharyngeal tumor is present, postcontrast MRI may offer the best tumor mapping while providing information about skull base marrow invasion, gross skull base bone destruction, and the presence of any intracranial tumor extension, be it by direct destruction of the skull base or via retrograde nerve extension. A gross assessment of any cervical nodal metastasis also can be obtained; however, full imaging coverage of all the nodal regions of the neck requires multiple sequences, is thus very time consuming, and is not as accurate as a postcontrast CT study. If information is needed regarding focal skull base or midface bone erosion, a CT study is needed.

If there is an oropharyngeal primary, postcontrast CT is usually the best initial examination. This study provides tumor mapping and excellent assessment of cervical nodal disease and identifies the presence of any mandibular or skull base bone erosion. If the tumor margins are not well seen, postcontrast MRI of that particular region often is helpful.

If there is an oral cavity tumor, most radiologists prefer CT as the initial examination. This study provides good tumor mapping, excellent nodal assessment, and evaluation of mandibular bone erosion. If the tumor is against the lingual cortex of the mandible, a Dentascan may best identify any focal bone invasion. In some cases, tumor mapping may not be ideal, and postcontrast MRI may provide better tumor visualization. MRI also may identify mandibular marrow tumor invasion better than can CT. However, if the patient has numerous dental amalgams, not only will image degradation artifact be present on CT scans, but there is also field distortion artifact on MRI. This distortion may become an almost insurmountable problem in the postoperative patient who has had mandibular reconstruction, where the stabilizing hardware often causes so much image distortion that the best CT and MRI examinations can give only limited information about the floor of the mouth and the anterior tongue.

If there is a question of tumor extension to the midface region, CT is the best initial study because it shows not only soft tissue tumor mass but also focal bone erosion. If a question persists regarding the distinction between tumor and sinonasal inflammatory disease, MRI can be useful in helping to resolve this specific query.

If there is a hypopharyngeal tumor, again, most radiologists prefer postcontrast CT as the initial study. This examination provides good tumor mapping, assessment of nodal disease, and an evaluation of laryngeal cartilage invasion. If a particular tumor margin is not well seen, postcontrast MRI may provide a better evaluation.

If the primary aim of the imaging study is to detect the presence of metastatic cervical nodal disease, postcontrast CT is the examination of choice. Because this is a major concern not only on the pretreatment evaluation but also on surveillance scans, CT is often the favored examination. If a particular soft tissue fullness is identified on such a CT study, a complementary postcontrast MRI may identify a dense fibrosis in a few cases. However, in most cases, MRI cannot better distinguish than CT between postoperative healing, the vascular type scar that occurs in the postoperative and postradiated neck, and tumor. Such difficult masses often come to CT-guided skinny needle aspiration biopsy.

In summary, today most radiologists prefer postcontrast CT as the initial imaging examination both in the pretreatment and the posttreatment patient with head and neck cancer. This in part reflects the fact that it is cheaper and faster than MRI. If a specific mass requires further definition, a postcontrast MRI study of that area is suggested. If there is a question of any intracranial tumor extension, postcontrast MRI is the examination of choice. With experience, the number of patients needing both CT and MRI studies decreases, the one exception being the skull base, where both CT and MRI are required to make the most informed imaging appraisal.

The basic imaging criterion to identify the presence of tumor on either CT or MRI is the identification of an abnormal mass. With cancer, the tumor margins infiltrate the adjacent fat planes and muscles, and in the case of oral cavity carcinoma, the jaw may be involved. Once the primary tumor is mapped, the presence of regional nodal metastases must be noted, and the entire upper aerodigestive tract should be scanned to identify the possible presence of a second primary tumor. With this information, today the clinician can formulate the most informed treatment plan possible.

The preoperative evaluation of the patient with cancer also requires assessment of regional nodal metastasis because the presence of pathological adenopathy is the single most important prognostic factor in these patients. The basic criteria used in the imaging evaluation of nodal disease are similar on CT and MRI. These criteria include nodal size, number, location, the presence of central necrosis, and extracapsular spread.

IMAGING ANATOMY

Whether CT or MRI is used in imaging of the neck, the basic concepts of imaging anatomy remain the same. A spatial approach is used to localize a mass or lesion within a defined space of the neck and then generate a differential diagnosis based on its appearance and location (**Table 56–2**). The spaces of the neck can be divided into the submandibular, sublingual, masticator, buccal, parotid, carotid sheath, parapharyngeal, prevertebral, visceral, retropharyngeal, and posterior cervical spaces.

The submandibular space is located inferior to the mandible, between the mylohyoid muscle and the hyoid

TABLE 56–2 SPATIAL APPROACH TO IMAGING OF THE NECK

Space	Contents	Common Pathology
Submandibular	Submandibular gland Lymph nodes Digastric muscle Facial artery and vein	Salivary gland tumors Abscesses Ranula Branchial cleft cysts Reactive lymph nodes Metastatic lymph nodes from squamous cell cancer oral cavity
Sublingual	Sublingual gland Lingual artery and vein Hyoglossus muscle Hypoglossal and lingual nerves Lymph nodes	Salivary gland tumors Abscesses Ranula Infections of odontogenic origin Squamous cell cancer
Masticator	Medial and lateral pterygoid muscles Temporalis muscle masseter muscle ramus of the mandible Cranial nerve V, third division Ramus of the mandible	Odontogenic abscess Masseteric hypertrophy Lymphoma Squamous cell cancer Rhabdomyosarcoma
Parotid	Parotid gland Facial nerve Retromandibular vein Periparotid and intraparotid lymph nodes Branches of external carotid artery	First branchial cleft cyst menangiomas Pleomorphic adenoma Mucoepidermoid carcinoma Adenoid cystic carcinoma Squamous cell carcinoma lymphoma Lymphoepithelial lesions Warthin's tumor
Carotid Sheath	Carotid artery Internal jugular vein Lymph nodes Cranial nerves IX, X, and XI	Reactive lymph nodes Metastatic lymph nodes Squamous cell carcinoma Lymphoma Paragangliomas Schwannomas Carotid artery aneurysm Carotid artery pseudoaneurysm Jugular vein thrombosis
Parapharyngeal	Fat Lymph nodes Minor salivary gland rests Internal maxillary artery	Minor salivary gland tumors Neurogenic tumors Abscess Tumor spread from adjacent spaces
Prevertebral	Prevertebral and paraspinal muscles Vertebral bodies Disk spaces	Osteomyelitis Diskitis Primary vertebral body tumors Metastasis Anterior osteophytes
Visceral	Nasopharynx Oropharynx Hypopharynx Larynx	Abscess Laryngoceles Diverticulum Goiters

TABLE 56–2 (*Continued*)

Space	Contents	Common Pathology
	Trachea Thyroid and parathyroid glands Esophagus Lymph nodes	Colloid cysts Adenomas Squamous cell cancer Thyroid cancer Esophageal cancer
Retropharyngeal	Fat Lymph nodes	Cellulitis Abscess Edema Suppurative adenitis Lymphoma Metastatic lymph nodes Squamous cell carcinoma
Posterior cervical	Lymph nodes Fat Spinal accessory nerve	Cystic hygroma Lipoma Reactive lymph nodes Lymphoma Metastatic lymph nodes Squamous cell carcinoma

bone. This space contains the submandibular gland, submandibular and submental lymph nodes, anterior belly of the digastric muscle, facial artery and vein, and fat. Most of the pathological processes in this space arise from either the submandibular gland or the lymph nodes. The most common lesion in adults is metastatic involvement of the lymph nodes from squamous cell cancer of the oral cavity. In children congenital lesions such as a branchial cleft cyst are more common. Other disease processes in this area include salivary gland tumors, abscesses, ranula, thyroglossal duct cyst, and dermoids. The submandibular duct can become obstructed and dilate, creating the appearance of a cystic mass (**Fig. 56–1**).

The sublingual space is located within the oral cavity superior to the mylohyoid muscle. It contains the hyoglossus muscle, hypoglossal and lingual nerves, lingual artery and vein, sublingual gland, and a portion of the submandibular gland and duct. It also contains lymph nodes. Common disease processes in this space include squamous cell cancer and infections of odontogenic origin. Other lesions include minor salivary gland tumors, ranula, thyroglossal duct cyst, and dermoids. If the hypoglossal nerve is injured, there will be loss of muscle volume on the affected side and fatty infiltration. This can simulate the appearance of an infiltrating tumor in the contralateral normal side.

The masticator space contains all four of the muscles of mastication: the medial and lateral pterygoid, masseter, and temporalis muscles. Also included in this space are the ramus of the mandible and the third division of the trigeminal nerve. This space extends superiorly to the

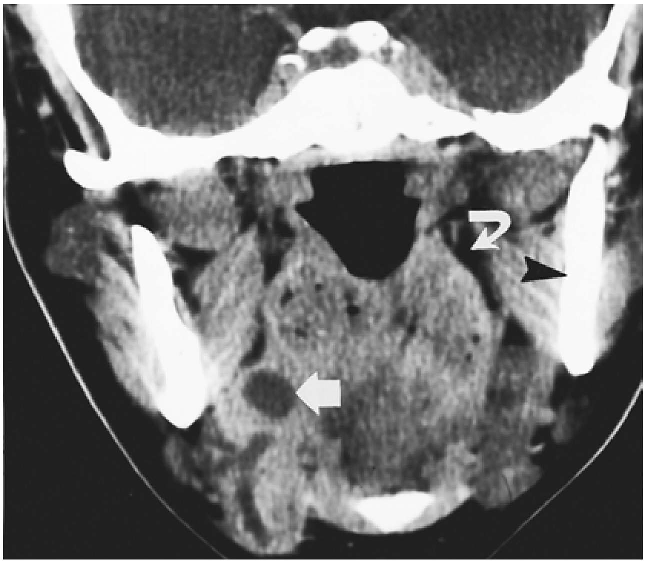


Figure 56–1 *Obstructed submandibular duct. Coronal computed tomographic scan of the oral cavity shows dilated submandibular duct within the right side of the floor of the mouth (straight arrow). The curved arrow points to fat within the parapharyngeal space. The arrowhead points to the mandible within the masticator space.*

skull base and temporal fossa. The most common lesion in this space is an odontogenic abscess. Neoplastic processes include sarcomas, lymphoma, and invasive squamous cell carcinoma. Neoplastic processes in the masticator space can spread via retrograde perineural extension along the trigeminal nerve into the intracranial compartment. In children, rhabdomyosarcoma also can occur in this area.

Masseteric hypertrophy can occur and should not be mistaken for a pathological mass. Masseteric hypertrophy presents as enlargement of the muscle, and the diagnosis is confirmed if this is bilateral or accompanied by enlargement of the other masticator muscles. If isolated to just one masseter muscle, the borders should be well defined, and a history of teeth grinding may be elicited. If there is still doubt as to the diagnosis, follow-up scans can be obtained.

The buccal space has no true fascia boundaries and can be involved with processes extending from the masticator space such as tumor or infection. The buccinator muscle and the superficial muscles of facial expression border the buccal space. It contains mostly fat as well as minor salivary glands, the facial artery and vein, and the parotid duct. Common disease processes in this space are abscess, cellulitis, squamous cell carcinoma, minor salivary tumors, and lipomas. An obstructed parotid duct can present as a palpated mass in this region.

The parotid space consists mainly of the parotid gland as well as portions of the facial nerve, retromandibular vein, and external carotid artery. There are also intraparotid and periparotid lymph nodes.

The parotid gland is split by the facial nerve into superficial and deep portions. The facial nerve itself is not well seen on imaging studies, but the retromandibular vein serves as a landmark for its course within the parotid gland. The most common lesion in this space is a pleomorphic adenoma. Malignant lesions include mucoepidermoid and adenoid cystic carcinoma as well as squamous cell carcinoma and lymphoma. Benign lesions include first branchial cleft cysts in adults and hemangiomas and lymphangiomas in children. In patients who are positive for human immunodeficiency virus (HIV), benign lymphoepithelial lesions can present as both solid and cystic lesions within the parotid, often associated with cervical adenopathy (**Fig. 56–2**).

Multiple masses within the parotid space mainly limit the differential diagnosis to Warthin's tumor, adenopathy (such as squamous cell carcinoma and lymphoma), and benign lymphoepithelial lesions.

The carotid sheath contains the carotid artery, internal jugular vein, lymph nodes, and cranial nerve (CN) IX, X, XI, and XII. It extends from the skull base to the aortic arch. There is a direct communication through the jugular fossa into the intracranial compartment, so tumors can extend above and below the skull base. The carotid sheath borders on the parotid space, parapharyngeal space, and retropharyngeal space. The most common pathology within the carotid sheath is enlarged lymph nodes, which can be from squamous cell carcinoma metastasis and non-Hodgkin's lymphoma. Primary tumors in this region include paragangliomas and schwannomas. Vascular lesions include jugular vein thrombosis and carotid artery aneurysm and pseudoaneurysm. On MRI, paragangliomas can be differentiated from schwannomas by the presence of flow voids within the mass. Flow voids may not be seen in lesions smaller than 2 cm.

The parapharyngeal space is bordered by the masticator and parotid spaces laterally, the carotid sheath

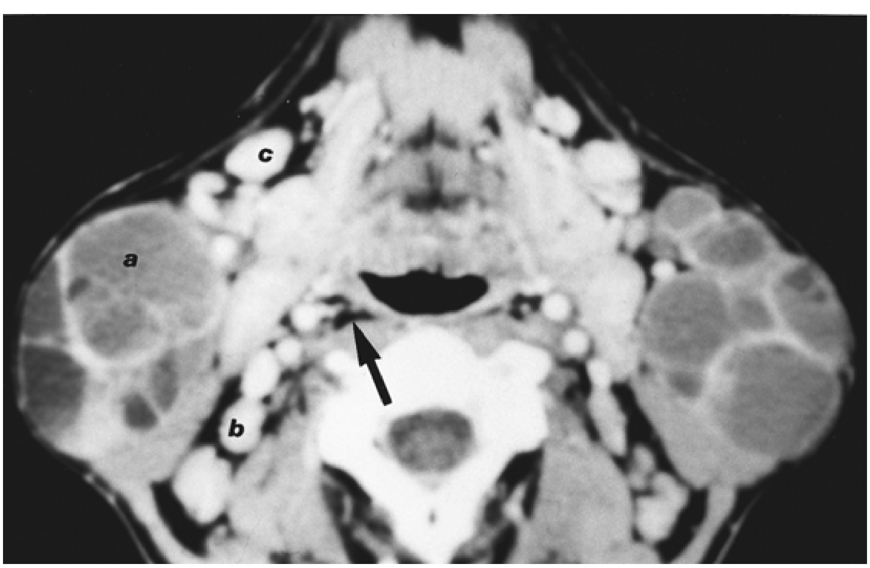


Figure 56–2 *Human immunodeficiency virus–related parotid disease. Axial computed tomographic scan of the neck shows (a) multiple cystic lymphoep-ithelial lesions within the parotid space. There are also (b) multiple enlarged spinal accessory chain lymph nodes in the posterior cervical space and (c) multiple enlarged internal jugular chain lymph nodes within the submandibular space. The straight arrow points to fat within the retropharyngeal space.*

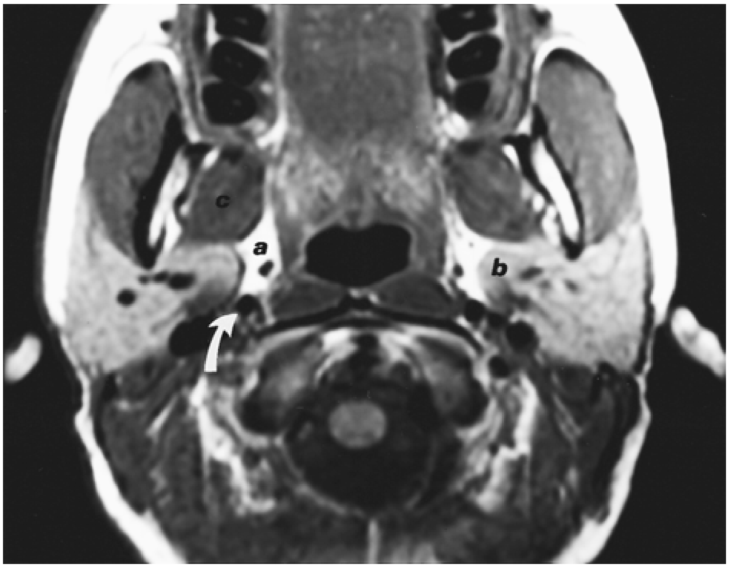


Figure 56–3 *Normal anatomy. Axial T1-weighted sequence from a magnetic resonance imaging scan of the neck. (a) The parapharyngeal space is bright in signal intensity because it contains fat. (b) The deep portion of the parotid space and (c) the pterygoid muscle within the masticator space are directly adjacent to the parapharyngeal space. The curved arrow points to the right internal carotid artery within the carotid sheath.*

posteriorly, and the submandibular space inferiorly (**Fig. 56–3**). The medial border is the pharyngeal mucosal surface, and it extends superiorly to the skull base. This space contains mostly fat as well as vascular structures such as the internal maxillary artery and ascending pharyngeal artery. There are also lymph nodes and minor salivary gland rests. The most common lesions in the parapharyngeal space are lesions spreading from the adjacent spaces. Squamous cell cancer can extend from the mucosal surface; sarcomas can come from the masticator space, salivary gland tumors from the parotid space. Abscesses can extend from the adjacent spaces such as odontogenic infection from the masticator space. Primary lesions arising in the parapharyngeal space are usually limited to minor salivary gland tumors and neurogenic tumors arising from branches of CN V.

It is important to differentiate lesions arising within the parapharyngeal space from lesions arising from the deep portion of the parotid gland. This most commonly occurs with pleomorphic adenomas. If the tumor is arising from the parotid space, the surgical approach will be different than from tumor in the parapharyngeal space to protect the facial nerve. If a lesion is surrounded by parotid tissue or extends into the superficial portion of the parotid gland, it most likely arises from the parotid gland. A lesion arising primarily within the parapharyngeal space may have a border of fat separating it from the parotid gland.

The prevertebral space contains the prevertebral and paraspinal muscles, vertebral bodies, disk spaces, and vertebral artery and vein. It lies directly posterior to the retropharyngeal space. The majority of disease in this region arises from the vertebral body or disk space. Thus usually there is an abnormal appearance of the vertebral bodies on imaging that confirms the presence of the lesion within the prevertebral space. The most common neoplastic processes are bony metastases. Primary lesions include chordoma and neurofibroma as well as primary vertebral body tumors. Inflammatory lesions are most commonly osteomylelitis and diskitis. Degenerative changes of the cervical spine can result in anterior osteophytes that can impinge upon the aerodigestive tract and cause dysphagia. In children, rhabdomyosarcoma and neuroblastoma can arise in this space.

The deep cervical fascia encloses the visceral space. In the suprahyoid neck it contains the nasopharynx and oropharynx. In the infrahyoid neck it contains the hypopharnyx, larynx, and trachea, as well as the thyroid and parathyroid glands. The esophagus, recurrent laryngeal nerve, and lymph nodes are also within this space. It extends from the skull base to the mediastinum. The most common malignancy in the suprahyoid portion is squamous cell carcinoma. Inflammatory lesions are usually odontogenic or tonsillar in origin.

The infrahyoid portion of the visceral space can be split into laryngeal, thyroid, parathyroid, and esophageal areas. Laryngeal lesions are located within the laryngeal cartilage, and the most common lesions are squamous cell carcinoma and laryngoceles. Esophageal lesions are centered on the esophagus, and the most common lesions are carcinoma and Zenker's diverticulum. Carcinoma of the esophagus usually spreads longitudinally before it invades into adjacent structures. The diagnosis is usually made by barium swallow or endoscopy. Cross-sectional imaging is used to evaluate for extension into surrounding tissues and nodal disease.

Lesions centered in the thyroid are usually of thyroid origin. The carotid artery is displaced laterally. Common benign lesions are goiters, colloid cysts, and adenomas. Malignant lesions include thyroid carcinoma, lymphoma, and metastases. CT or MRI cannot reliably differentiate malignant from benign thyroid lesions. Findings that suggest malignancy are extracapsular spread, cartilaginous or bony destruction, and recurrent laryngeal nerve dysfunction. Nodal disease is also an indicator of probable malignancy. Calcifications and cystic areas within the tumor are not useful in predicting benign form malignant disease. In patients with suspected thyroid malignancy, MRI is favored over contrast-enhanced CT because the iodinated contrast used in CT can delay iodine 131 therapy.

In evaluating parathyroid disease, ultrasound and radionuclide studies are the preferred imaging modality. Ultrasound is recommended for evaluation of localized parathyroid pathology, and radionuclide studies with thallium and technetium are necessary for finding ectopic parathyroid adenomas. CT and MRI should be used after the radionuclide study to define the location of the ectopic adenoma anatomically. MRI is more specific in identifying an adenoma, with the adenoma seen as bright signal on T2-weighted sequences.

The retropharyngeal space is directly posterior to the visceral space and anterior to the prevertebral space. The retropharyngeal space extends from the skull base to the diaphragm. The posterior portion of the retropharyngeal space is called the danger space because head and neck infections can be transmitted inferiorly into the posterior mediastinum. The retropharyngeal space contains mainly lymph nodes and fat. The lymph nodes in this space are also called the nodes of Rouvière and are split into medial and lateral groups. Lymph nodes are seen only down to the level of hyoid bone; below this level this space contains only fat. The most common lesions seen in this space originate from the lymph nodes and include metastatic adenopathy and suppurative adenitis as well as lymphoma. Other processes seen in this space are inflammatory and include cellulitis and abscess. Edema also may be seen in this space, especially in patients who have received radiation therapy. Infection and neoplastic processes can spread into the retropharyngeal space from adjacent spaces such as squamous cell carcinoma from the visceral space and diskitis and osteomyelitis from the prevertebral space.

The posterior cervical space is located posterior and lateral to the carotid sheath. It contains fat, the spinal accessory nerve, and the spinal accessory lymph node chain. Lesions in this space are usually nodal in origin. The most common malignant lesion is squamous cell metastatic disease. The primary site of disease is usually in the nasopharynx or oropharynx. Benign lesions include reactive and suppurative adenopathy as well as lipomas. In children the posterior cervical space is a common site for cystic hygromas and lymphangiomas.

CYSTIC MASSES OF THE NECK

Cystic masses of the neck often have a characteristic appearance and location that allow a specific diagnosis to be made based on the imaging findings (**Table 56–3**). A cystic mass will appear low density on CT scanning with no central enhancement. MRI will show cystic lesions as low signal intensity (dark) on T1-weighted sequences and high signal intensity (bright) on T2-weighted sequences. On postcontrast T1-weighted sequences cystic lesions remain low signal intensity.

Cystic lesions of the neck are often congenital. Common congenital lesions include thyroglossal duct cyst, branchial cleft cyst, cystic hygroma, and dermoid. Thyroglossal duct cysts arise from remnants of thyroid tissue along the course of the thyroglossal duct. The thyroglossal duct is a tract of epithelial tissue extending from the base of the tongue to the normal position of the thyroid gland. Most thyroglossal duct cysts are at the level of the hyoid or below. The cyst can be midline or slightly off midline. The more inferior the cyst is located, the more likely it is to be off midline. There may be a tract leading toward the hyoid bone, and the cyst is usually within the strap muscles of the neck.

Branchial cleft cysts can occur at various levels in the neck. First branchial cleft cysts are seen at the level of the parotid gland. Second branchial cleft cysts are the most common and occur at the level of the angle of the mandible, usually along the anterior border of the sternocleidomastoid muscle. Branchial cleft cysts, although congenital, can increase in size acutely, often after an upper respiratory infection. A component of the second

TABLE 56–3 COMMON CYSTIC LESIONS OF THE NECK

Thyroglossal duct cyst
Branchial cleft cyst
Cystic hygroma
Dermoid
Abscess
Jugular vein thrombosis
Ranula
Laryngocele
Infected lymph nodes
Metastatic lymph node (usually thyroid primary)

branchial cleft cyst may lead toward the space between the internal and external carotid arteries. Third branchial cleft cysts are rare and are seen in the posterior cervical space.

Cystic hygromas usually present before the age of 2 years, although they can present in the second and third decade of life. Cystic hygromas and lymphangiomas are a congenital malformation of lymphatic channels. They are most commonly seen in the posterior cervical space but can traverse several spaces of the neck. When imaged, these lesions are often multiloculated masses that surround normal structures. There may be vascular components of the lesion, which do enhance on postcontrast studies.

Dermoids and epidermoids present as slow-growing masses in the oral cavity. Epidermoids are fluid density, whereas dermoids contain mixed elements that may be fat density (dark) on CT and fat intensity (bright) on T1-weighted MRI sequences. Epidermoids tend to be in the sublingual space, often along the midline, and dermoids are often in the submandibular space.

Cystic lesions in the neck can be inflammatory in nature. Abscesses will present as a cystic mass that may be loculated and often with a thick enhancing wall. There will usually be surrounding inflammatory changes such as skin thickening and irregularity of the surrounding fat. Possible causes of an abscess include dental infection, pharyngitis, osteomyelitis, and salivary gland calculus disease. Infection within a lymph node can result in suppurative adenopathy. Lymph nodes will appear with a low-density center on CT. There can be single or multiple lymph nodes involved. Suppurative adenopathy can have the same appearance as pathological lymph nodes with metastatic disease and central necrosis. Biopsy may be required to distinguish between these.

Other inflammatory lesions that can present as cystic masses are retention cysts associated with mucus-producing glands. The most common of these is a retention cyst of the sublingual gland, which is called a ranula. A ranula can be contained just within the sublingual space and is called a simple ranula. If a simple ranula ruptures and extends into the submandibular space, it is called a plunging or diving ranula.

Other lesions that can present as cystic lesions are jugular vein thrombosis and laryngoceles. Visceral space cystic lesions can be of thyroid origin, such as colloid cysts and multinodular goiters. Less common visceral space lesions are parathyroid and thymic cysts. Metastatic disease within lymph nodes can have a cystic appearance, especially if the primary tumor is of thyroid origin.

IMAGING NODAL DISEASE

When looking for possible nodal pathology, there are several different methods used to measure the nodal size of a homogeneous sharply delineated node. All of the methods have an 80 to 85% accuracy rate, so that it matters little which method is used. The most commonly used method measures the greatest longitudinal nodal length. Less commonly utilized methods measure the maximal transverse nodal size, and another method involves making a ratio of these two measurements. In the floor of the mouth and for the jugulodigastric lymph nodes, any node larger than 1.5 cm in longest dimension is considered pathological. For all other nodal groups, any node larger than 1.0 cm in longest dimension is considered pathological. If a lymph node is larger than these guidelines, there is an ~80% incidence of metastatic involvement on histologic analysis. For such homogeneous sharply defined nodes that are normal in size but at risk for metastasis due to location, ~20% may contain microscopic tumor. Thus the clustering nodes in an at-risk area for tumor spread should be considered a finding suspicious of tumor, despite the otherwise normal imaging appearance of these nodes. One of the major limitations of clinical evaluation, CT, and MRI is the inability to identify metastasis in homogeneous nonenlarged nodes, and it is to this issue that future imaging research is addressed.

At present, the imaging gold standard criterion for nodal metastasis is the presence of central nodal necrosis, and this is best identified on either contrast-enhanced CT or fat-suppressed contrast-enhanced T1-weighted MRI (**Table 56–4**). The central "necrosis" appears as a central region of low attenuation on CT or low to intermediate signal intensity on the T1-weighted MRI scan. This central area actually represents both tumor cells and necrosis, and this causes some slight differences on CT and MRI. Specifically, these areas on noncontrast T2-weighted MRI appear nonhomogeneous, with sites of both high and low signal intensity. However, the areas of high T2-weighted signal intensity do not correspond precisely to the regions of low density seen on the CT scans. This reflects the fact that

TABLE 56–4 IMAGING FINDINGS OF METASTATIC LYMPH NODES

Central necrosis (this is the most specific sign)
Length greater than 1.0 cm (1.5 cm for jugulodigastric nodes)
Clustering of lymph nodes
Extracapsular spread

the regions of high T2-weighted MRI signal intensity represent sites of necrosis, and the low-density region on CT represents both necrosis and tumor. Regardless of size, nodes with central necrosis are always considered pathological.

Extracapsular tumor spread is identified on CT and MRI by an ill-defined peripheral nodal enhancement with loss of the normal fat planes between the node and adjacent structures. This imaging appearance can be seen with acutely inflamed nodes, recently irradiated nodes, and nodes in a recent surgical bed. Thus these conditions must be clinically correlated to maximize the reliability of the imaging study. Extracapsular tumor spread is presently more reliably identified on CT than MRI.

The advantages of CT and MRI over clinical examination include the identification of central "necrosis" in a normal-sized node. CT and MRI can better visualize retropharyngeal nodes and high internal jugular nodes, as well as adenopathy deep to the sternocleidomastoid muscle and in the tracheoesophageal groove, areas difficult to evaluate by palpation. Most of the time, nodal staging is preferentially done with CT because studies have shown that CT is more accurate than either clinical examination or MRI. This reflects the fact that at present, central nodal necrosis is better seen on CT. However, with future developments in MRI contrast agents, MRI may play a more important role in nodal staging. Other possible modalities that may play a role in the future for nodal staging include positron emission tomography.

POSTOPERATIVE IMAGING

Both clinically and radiographically, the evaluation of the postoperative patient is usually more difficult than the preoperative assessment. The clinical examination can be limited by thickening of the skin due to prior irradiation and/or postoperative edema from venous compromise, whereas the bulk of a flap makes deep palpation difficult. Similarly, cross-sectional imaging of the postoperative patient is challenging because of the loss of the normal symmetry of the neck structures and the resulting altered anatomy from the wide array of reconstructive flaps presently available to the head and neck surgeon. In addition, postoperative and postirradiation edema can obscure fat planes, simulating tumor infiltration. If the radiologist is unaware of the type of surgery, whether or not the patient was irradiated, and when these events occurred in relationship to the imaging study, the CT and MRI examinations can be confusing and prone to misinterpretation. The imaging can be made further difficult because of artifact from reconstruction plates. Whether this hardware is made from titanium or some other alloy, in the best of circumstances some field distortion on MRI and degradation artifact on CT remain imaging problems. Although the current hardware is not affected by either the magnetic field of MRI or the photons of CT, the artifact problem remains a major issue because the anterior tongue and floor of the mouth often cannot be well imaged in these patients. In addition, if the patient was irradiated, fibrosis of muscles and soft tissue planes can, at times, be confused with tumor. Lastly, the granulation tissue and the "vascularized scar" that occur in the neck after surgery and irradiation have the same imaging characteristics as tumor. Thus the cooperation between clinician and imager, which plays an essential role in the preoperative imaging of the patient with oral cancer, is even more important in imaging the postoperative patient.

WHEN TO IMAGE

The postoperative patient can be imaged in three different situations. The first is for the evaluation of an immediate postoperative complication that occurs within the first or second postoperative week. The second situation is to establish a base line CT or MRI examination, a study to which further imaging studies will be compared. This base line examination is usually obtained between 4 and 8 weeks postsurgery. The third situation is the postoperative surveillance of the patient; this should occur according to a protocol, usually for a period of at least 3 to 5 years after initial treatment.

During the first few days after surgery, the main concern is postoperative infection, hemorrhage, and, in the appropriate case, flap viability. This last complication is best assessed clinically. However, imaging can be of great assistance when evaluating the presence of either infection or hemorrhage. Specifically in the case of infection, the primary problem is the presence of an abscess, and this is well seen on imaging. In general, CT is the preferred modality during this immediate postoperative period because the patient is often unable to remain motionless long enough for a good diagnostic quality MRI study to be performed. In addition, many of these patients are medically unstable and require monitoring devices, many of which cannot be used within the high magnetic field of the MRI unit. Special nonferromagnetic life support systems are necessary, and they are often not available at the time of the examination.

The second time frame for imaging the postoperative patient occurs between 4 and 8 weeks after surgery and/or radiation therapy, when a base line imaging study should be obtained. The desire is to image the patient when most of the postoperative edema and hemorrhage have resolved but before any recurrent tumor is present. In general, in a patient with pathologically tumor-free surgical margins, it is rare for a recurrence to occur at 4 to 8 weeks following surgery. The importance of such a base line examination cannot be overemphasized because the often complex postoperative anatomical changes must be documented when no tumor is present. The follow-up studies can then be most accurately evaluated by comparison to this base line study. In some patients, especially those who have had postoperative radiation, these edematous type changes can persist for 12 to 18 months after therapy. Although currently both CT and MRI play an important and complementary role in following these patients, CT remains the preferred initial modality.

The surveillance of the postoperative patient is usually based on a protocol that takes into account the statistics on time of recurrences, which show that most recurrences develop within 2 years following surgery and that most deaths from disease occur within 3 years of surgery. Thus the suggested protocol is to have clinical and imaging surveillance examinations at 6-month intervals for the first 2 or 3 years and then at least once a year for the next 2 or 3 years. Whenever a clinical problem develops, such as pain, neurological deficit, or a mass, the patient is scanned at that time to best detect a potential early recurrence. Both MRI and CT can be used to follow these patients. CT offers the quickest and most reliable way to detect nodal recurrence, whereas a recent study (Lell et al., 2000) has shown that MRI may be better than CT for differentiating recurrent disease from radiation fibrosis. The problem of artifact from the reconstruction hardware has already been mentioned. On CT, images through these regions can be viewed at high window settings to minimize the visualization of the artifact. However, the soft tissues become "grayed" with low contrast, decreasing the effectiveness of the CT study. Although some physicians believe that MRI is unaffected by such artifact, field distortion from the hardware occurs in most patients and causes a "black region," in which all detail is lost. Often this region of field distortion is smaller than the area affected by artifact on CT, and a combination of CT and MRI may better evaluate the postsurgical bed.

THE IMAGING APPEARANCE OF RECURRENT DISEASE

The imaging criteria to evaluate the presence of recurrent tumor apply regardless of the precise location of the recurrence within the head and neck. The essential finding is the presence of a soft tissue mass, usually with infiltrative margins. This mass may be homogeneous or partially necrotic. Specifically, the appearance of any mass on a surveillance scan, not identified on an earlier study, is a recurrence until proven otherwise. Postoperative physiological changes only diminish in size or remain constant in appearance on subsequent studies. One of the diagnostic problems is the distinction between recurrent tumor and the presence of vascularized scar, a tissue whose imaging characteristics, including the degree of enhancement, on both CT and MRI are indistinguishable from tumor. In some cases, CT-guided skinny needle aspiration of the tissue may be the only way to try and resolve the issue. On CT, tumor has attenuation similar to that of muscle, reflecting its cellular composition. If soft tissue fullness is present, but of an attenuation less than that of muscle, the mass is almost always edematous tissue. Similarly, the attenuation of tissue that infiltrates fat planes is important because tissue that is as dense as muscle is usually a recurrence, whereas a region less dense than muscle almost always represents fluid and edema. Such fluid density material may persist for years and is often identified in the retropharyngeal space in postoperative and irradiated patients.

Most recurrences occur at surgical margins, be they soft tissue or osseous. Surgical bone margins are usually sharp and clearly delineated. Any bone margin that has irregular, infiltrative edges must be considered either tumor recurrence or osteomyelitis. The latter is especially prone to occur in the postirradiated patient. In addition, the edentulous jaw tends to become demineralized, often giving the imaging suggestion of being infiltrated by tumor. This appearance is further complicated in that the normal high T1-weighted MRI signal intensity of fatty mandibular marrow is diminished by both tumor and postradiation fibrosis.

If the pharynx was partially resected and a free flap utilized, or the patient had a jejunal autograft for reconstruction of the pharynx, recurrences can present as extraluminal or intraluminal masses, often located at the upper anastomotic site. This may cause narrowing or an eccentric lumen of the pharyngeal airway with proximal pharyngeal dilatation. However, redundancy of the flap, edema, and scarring can all mimic a recurrent mass. Further complicating the imaging in patients with a tracheostomy is the fact that due to preferential breathing through a tracheostomy, the pharynx tends to collapse.

Most patients will have undergone some form of neck dissection, the most extensive of which being the radical neck dissection. The imaging findings of a radical neck dissection include the absence of the ipsilateral sternocleidomastoid muscle, internal jugular vein, and submandibular gland. Atrophy of the trapezius muscle develops secondary to sacrifice of the spinal accessory nerve, and there usually is compensatory hypertrophy of the ipsilateral levator scapulae. There is also flattening of the contour of the neck, and the carotid artery often lies directly deep to the skin.

Modified neck dissections are less extensive, with preservation of one or more of the sternocleidomastoid muscle, internal jugular vein, and spinal accessory nerve. The imaging findings of these modified radical neck dissections may be subtle, including slight flattening of the contour of the neck and lack of sharpness of the local tissue planes. There also may be minimal thickening of the skin at the scar site. Alternatively, in cases with extensive disease, there may be removal of additional nodal groups and/or nonlymphatic structures, such as the platysma, strap muscles, trapezius muscle, thyroid gland, CN IX and X, and carotid artery.

If the surgical defect is too large for a primary closure, one of several surgical flaps can be used in the reconstruction. The most commonly used flap is the myocutaneous flap, which consists of skin, subcutaneous tissue, and muscle. Typical sites of origin include the pectoralis and trapezius muscles, and the vascular pedicle of the flap can always be traced to the muscle of origin on the images. Because the nerve supply to the muscle is sacrificed, the muscle atrophies, eventually resulting in fatty muscular degeneration. This commonly results in the CT and MRI appearance of the flap as a thickened area of fatty tissue with overlying normal skin, within which there are streaks of soft tissue density that represent the residual degenerated muscle fibers.

The MRI appearance of these flaps reflects the signal characteristics of its components. Thus the fatty bulk of the flap has high signal intensity on T1-weighted sequences, which fades to intermediate signal intensity on the T2-weighted sequences. Any remaining muscle has an intermediate signal intensity on T1-weighted images and a lower signal on T2-weighted scans. Recent studies (Chong et al., 2001) have indicated that enhancement characteristics of the surgical flap are variable, with some cases having no enhancement, however, the remaining muscular component may demonstrate significant enhancement.

After a craniofacial resection, a temporalis free flap is used to close the anterior cranial fossa floor defect. This graft is eventually fibroses. On postcontrast CT, the graft usually is of muscle attenuation and enhances slightly. On MRI, the graft also enhances until it becomes completely fibrosed, at which time it has low signal intensity similar to that of the adjacent bone. In these patients, MRI shows the dural surface of the graft better that CT; however, the lateral margins of the graft will be better identified on CT.

The radiologist also has to be able to identify other commonly used flaps. Thus, especially in the reconstruction of the floor of the mouth, a free flap is best identified on MRI and CT as replacement of the normally anatomic structures by a bulky, mainly fatty, soft tissue mass. Unlike the myocutaneous flap, which has a vascular pedicle that is easily identified on CT and MRI, the free flap has no rotated local muscle and vascular pedicle. The osteomyocutaneous flap is easily identified on imaging of the postoperative mandible by the inclusion of bone, often from the iliac crest, within the myocutaneous flap.

It is also imperative in the postoperative imaging of patients with head and neck cancer to evaluate for metastatic nodal disease, and postoperative scans should always include the entire neck to assess the nodal chains. The criteria used to evaluate adenopathy in the postoperative neck are the same as those already discussed in the preoperative patient.

Although gross tumor recurrence does not occur within the first postoperative week, there are several conditions that can mimic the appearance of tumor recurrence on imaging. In these cases, clinical correlation usually readily clarifies the problem, and the radiologist less familiar with head and neck imaging of the postoperative patient may need to be counseled by the clinician. Thus, although a hematoma in the postoperative bed can simulate the appearance of a necrotic node on imaging studies, the hematoma presents within the first postoperative week when tumor recurrence is highly unlikely. Similarly, a postoperative lymphocele could present on imaging studies as a nodular mass; however, this lesion also usually occurs in the immediate postoperative period, and, on CT, it has an attenuation less than muscle. Another imaging imitator of pathological adenopathy is an abscess. However, the clinical presentation of pain, fever, and elevated white blood count usually identifies this entity. By correlating the imaging findings with the clinical information, an imaging misdiagnosis of recurrence usually can be avoided.

However, after several months, there will be some situations that even after correlating the clinical history and imaging findings and making comparisons with old films, the distinction between recurrent disease and postoperative change remains unclear. One option available in these circumstances is a CT-guided fine-needle aspiration. The suspicious mass is localized on an initial scout CT scan, and a percutaneous path to the lesion is chosen that avoids the major neurovascular structures. A 22-gauge Chiba needle is advanced into the lesion along this path, and CT scans are performed to help confirm the needle path. After the CT confirmation that the needle tip is within the suspicious lesion, biopsies are taken. These biopsies are best performed with a cytologist present who can confirm the adequacy of specimens. Several passes are usually required to attain adequate samples. The postoperative patient with a surgical flap usually tolerates this procedure well because the flap tends to have little or no sensation.

At present CT is used as the major imaging modality for guidance of needle biopsies. However, there are MRI-compatible biopsy needles available, and with advances in software, the time for scanning in MRI is constantly decreasing. In the near future, real-time MRI and CT scanning will be available, and so-called CT and magnetic imaging fluoroscopy will allow the radiologist to watch the advancement of the needle into the lesion and markedly decrease the time needed for imaging-guided biopsies. Another future role of MRI in the postoperative evaluation of patients with head and neck cancer is magnetic resonance spectroscopy. This technique will allow assessment of suspicious areas for their actual biochemical makeup, rather than their appearance. This will increase the specificity of diagnosis of disease recurrence and possibly avoid invasive procedures.

As already mentioned, the most effective postoperative imaging of the patient with head and neck cancer depends on a two-way dialogue between the clinician and the radiologist. The postoperative scans cannot be optimally interpreted without knowledge of the prior surgical procedures performed and the patient's current clinical status. The aim of such clinicoradiographic cooperative surveillance is to detect, as early as possible, occult recurrent disease and thus improve the patient's chances of survival.

Suggested Readings

Chong J, Chen LL, Langstein H, et al. MR imaging of the muscular component of mucocutaneous flaps in the head and neck. *American Journal of Neuroradiology* 2001; 22:170–174.

Harnsberger HR. Handbook of Head and Neck Imaging. 2nd ed. St. Louis: CV Mosby; 1995

Lell M, Baum U, Gress A, et al. Head and neck tumors: imaging recurrent tumor and post-therapeutic changes with CT and MRI. *European Journal of Radiology* 2000; 33:239–247.

Shankar L, Azra Khan A, Cheung G, eds. Head and Neck Imaging. New York: McGraw-Hill Professional Publishing; 1998

Som PM, Curtin HD, eds. Head and Neck Imaging. 4th ed. St. Louis: CV Mosby; 2002

Yousem DM. Case Review: Head and Neck Imaging. St. Louis: CV Mosby; 1998

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. Which of the following is not a contraindication for a patient to undergo magnetic resonance imaging (MRI) scanning?
   - A. Titanium reconstruction plates
   - B. Cardiac pacemaker
   - C. Intracranial aneurysm clip
   - D. Cochlear implant
2. Somatostatin analogue imaging is useful for evaluating which type of tumor?
   - A. Squamous cell carcinoma
   - B. Paraganglioma
   - C. Lymphoma
   - D. Schwannoma
3. The best study to evaluate for metastatic nodal disease in the neck is
   - A. Ultrasound
   - B. MRI scan
   - C. Computed tomographic (CT) scan
   - D. Plain films
4. The most specific sign of metastatic disease within a lymph node is
   - A. Central necrosis
   - B. Oblong shape
   - C. Enhancement
   - D. High signal on T1-weighted image

Chapter 57
THE AGING FACE

Ivan Wayne and Brian Jewett

Facial aging is a complex phenomenon involving multiple factors. Almost as soon as one's face reaches a mature size it begins to undergo changes associated with aging. The speed and extent of this process are determined by the interplay of external factors and genetic determinants. Superimposed on the physical process is the psychological response of the patient. Individuals will cope with aging differentially according to their sense of self, social support systems, and societal standards of beauty. Some will seek medical assistance to learn more about the prevention and treatment of changes associated with the aging face.

ANATOMY AND PHYSIOLOGY

The facial changes that occur with aging are well established and include changes in the superficial soft tissues as well as the underlying skeletal support of the face. The face can be divided into three layers. The skin is the facial covering, which consists of epidermis, dermis, pilosebaceous units, and a layer of subcutaneous fat containing blood vessels. Facial soft tissues make up of various fat pads, muscles, and connective tissues make up the second layer. An important component of the second layer is the superficial musculoaponeurotic system (SMAS), which consists of an interconnected system of musculature and supporting ligaments that facilitate facial movements. The foundation of the face is the bony skull and facial bones. Each of these facial structures undergoes changes as the body ages.

With aging, intrinsic changes in the facial skin occur that include reduction in skin cellular and protein components, epidermal thinning with retraction of rete pegs, loss of moisture, and pigmentary changes secondary to alterations in the number and distribution of melanocytes. Within the dermis and subcutaneous tissues, elastic filaments become thin and fragmented, resulting in increased skin laxity.

TABLE 57–1 GLOGAU CLASSIFICATION

Type	Wrinkling	Keratosis
I	None	None
II	Present with motion	Unlikely
III	Present at rest	Visable actinic keratosis and telangiectasias
IV	Wrinkles throughout	Actinic keratosis; often prior skin cancer

TABLE 57–2 FITZPATRICK SKIN TYPES

Type	Color	Characteristics
I	White	Always burns/never tans
II	White	Usually burns/rarely tans
III	White to olive	Mildly burns/average tans
IV	Light brown	Rarely burns/usually tans
V	Dark brown	Rarely burns/always tans
VI	Black	Never burns

Given that the face is exposed to the damaging effects of sunlight and wind, textural changes are often noted in the face earlier than other parts of the body. Environmental factors such as ultraviolet (UV) radiation cause degradation of collagen and elastic fibers, leading to solar elastosis, which is characterized by thickened, poorly organized elastic tissue. Solar injury to cellular deoxyribonucleic acid (DNA) can lead to atypia and carcinoma. Photodamaged epidermis shows increasing cellular dysplasias, which are probable precursors to actinic keratoses and skin malignancies. Melanin exposed to cumulatively large amounts of solar exposure can undergo hypertrophy and clumping, leading to pigmentary dyschromias. Sun damage and cigarette smoking can accelerate a reduction of vascular supply to the skin.

In addition to changes in the overlying facial covering, facial aging is characterized by changes in the underlying soft tissues and osseous structural framework. Fat atrophy, combined with the gravitational redistribution of soft tissues and loss of bone in areas of bony prominence, results in facial contour changes. The fat pads of the face, including the buccal, malar, and suborbicularis oculi fat pads, are intertwined with the supporting ligaments and muscles of the face. As these support structures weaken and relax, the persistent action of gravity leads to the descent of these facial soft tissues. As the facial skeleton resorbs, soft tissues become ptotic, and the effective skin surface area increases, facial wrinkles and creases become more prominent. The Glogau scale is used to describe the changes associated with the aging face (**Table 57–1**).

PREOPERATIVE EVALUATION

General Assessment

Patients must be assessed for comorbidities that increase the risk of perioperative complications. Diseases that adversely affect cutaneous blood supply or the coagulation process are of particular importance. Diabetes, alcoholism, and tobacco addiction can complicate healing wounds. Each patient undergoing surgical procedures should have a thorough preoperative medical evaluation by the surgeon and anesthesiologist.

One of the most important areas for the surgeon to address is patient motivation. In general, the best patients are those who have given serious consideration to undergoing cosmetic surgery and wish to have their outward appearance match their inner vitality and sense of self. Patients who have a greater risk of postoperative dissatisfaction are those who seek facial rejuvenation for ulterior motives, such as regaining the attention of a spouse or significant other.

An overall facial examination should be performed, noting any asymmetries, contour abnormalities, and surgical or traumatic scars. The skin should be inspected for dysplastic lesions, malignancies, and inflammatory conditions. The patient's skin type also should be assessed to anticipate the patient's response to potential treatment modalities, including skin resurfacing (**Table 57–2**).

Forehead Region

Aging of the forehead complex can create significant aesthetic and functional problems for the patient. Relaxation of the forehead skin and supportive ligaments allows the brow to descend, with gravity causing brow hooding. This deformity, often in combination with redundancy of upper eyelid skin, can obscure the superior visual fields. The patient will try to compensate with frontalis muscle hyperactivity to lift the eyebrows, causing horizontal creases, muscle fatigue, and possibly even headaches. Patients with ptotic brow tissues should be questioned about these symptoms, and objective testing should be obtained to document visual field deficits when indicated. The youthful female brow contour and position are described as being arched above the supraorbital rim, with the apex of the arc located at a point somewhere between the lateral limbus and lateral canthus. The youthful male brow should rest at the supraorbital rim, with a more linear, less arched path.

It is easy to be deceived when evaluating brow position. Frontalis contraction can make a ptotic brow appear normal; however, this is often a temporary position, with the brow descending as the frontalis fatigues after prolonged contraction during the course of the day. These patients often report that their brow feels heavy and that their visual complaints worsen in the evening. To determine resting brow position in these patients, have them close their eyes for several seconds to allow the frontalis muscle to relax. The brow is then fixed by the examiner's hand being placed against the forehead, and the patient is asked to open the eyes. The relaxed position of the brow is then determined, and the amount of correction needed to bring the brow to the correct location is measured. Normal distance from the upper lid margin to the central inferior brow is between 10 and 15 mm. Failing to recognize brow ptosis may lead to inappropriate skin excision during an upper lid blepharoplasty. Finally, symmetry of the frontalis muscle and integrity of the frontal branch of the facial nerve should be noted, and any asymmetry of brow position should be pointed out to the patient. The forehead should also be examined for the presence of vertical and horizontal rhytids.

Periocular Region

Changes in the aging forehead are often associated with changes in the periorbital region. As the periorbital complex ages, several changes may become apparent, including redundancy of upper lid skin, bulging of orbital fat, and wrinkling of periorbital skin. These findings, often combined with brow ptosis, make the eyes appear smaller, fatigued, and sometimes angry in appearance. In addition to the aesthetic consequences, these changes can have functional significance because redundant upper lid skin can impair the superior visual field. Changes also occur in the lower eyelid. Descent of the cheek soft tissue and bulging of lower eyelid fat interrupt the smooth, youthful transition from eyelid to cheek. In addition, lower eyelid laxity increases with aging. There are several ophthalmologic terms that are important to understand when evaluating patients.

Blepharochalasis is an unusual condition involving recurrent episodes of angioneurotic edema that result in relaxation and redundancy of eyelid skin. This is thought to be more common in young females.

- *Dermatochalasis* is characterized by acquired eyelid skin laxity secondary to actinic damage and intrinsic changes in eyelid skin associated with aging.
- *Steatoblepharon* occurs as the orbital septum weakens, allowing pseudoherniation of orbital fat. This condition is more commonly referred to as pseudo fat herniation.
- *Blepharoptosis* indicates malposition of the upper eyelid, often secondary to dehiscence of the levator aponeurosis, allowing a lower than normal lid position.
- *Lagophthalmos* is incomplete closure of the upper lid, allowing possible exposure of the cornea and sclera.
- *Ectropion* presents with lid margin retraction away from the globe and can be secondary to loss of eyelid tissue, scarring, or weakening of eyelid structural support.

Patients considering periocular facial rejuvenation should be evaluated for concomitant ophthalmologic conditions that may adversely affect patient outcome. Inquiring about a history of dry eyes, thyroid disease, and glaucoma is important. Some surgeons require all patients undergoing blepharoplasty to have an ophthalmologic consultation and clearance. Many surgeons will not operate on patients with dry eyes because of the risk of worsening the condition, especially if any postoperative lagophthalmos develops. Periocular complications associated with Graves' disease can be exacerbated by blepharoplasty.

The physical exam should include assessment of visual acuity in each eye, determination of ocular motility, and evaluation for dry eyes. The shape and height of the supratarsal lid crease also should be evaluated. Normally, the position of the upper lid crease to lid margin is 9 to 11 mm, and levator dehiscence may result in malposition of the upper eyelid crease. In addition, the upper eyelid is usually positioned at the superior aspect of the iris on frontal gaze. If the lid is closer to the pupil, blepharoptosis is present, and levator dehiscence should be considered. Levator function should be assessed by measuring upper lid excursion because patients with normal lid position may have impaired levator function. The amount of upper lid excursion is measured by holding a ruler in front of the closed lid and having the patient open the eyes while looking up. If this distance is less than 12 mm, an ophthalmologic consultation should be sought to elucidate the cause.

Upper eyelid fat pads are located centrally and medially, and the lacrimal gland is located laterally. Pseudoherniation of fat as well as ptosis of the lacrimal gland should be assessed. Lower eyelid fat pads are located medially, centrally, and laterally. Pseudoherniation of the lower orbital fat pads may be evident on primary gaze, but it is accentuated by having the patient look up.

The tone of the lower lid also should be examined. Lid distraction test involves grasping the lid between the finger and thumb and retracting from the globe. More than 10 mm maximum distraction indicates lid laxity. In addition, release of the distracted lower eyelid should produce a brisk return of the lid to its normal position (snap test). A delayed return or one that requires blinking reveals poor lid tone. If there is significant fine wrinkling in the periorbital skin, this should be discussed with the patient, and an appropriate resurfacing procedure should be planned, as blepharoplasty will not fully eliminate this problem.

Midface Region

The appearance of the midface and that of the lower eyelid are closely related. In the youthful face, the lower eyelid is concave in profile and gradually blends into the convexity of the midface–cheek complex without noticeable interruption. Important structures in the midface region include the malar, buccal, and suborbicularis fat pads. These fat pads become ptotic with age, resulting in tear trough deformities, prominent nasolabial folds, and a visible separation between the lower eyelid and midface region. A visible depression at the inferior orbital rim may develop secondary to soft tissue atrophy. In addition, descent of the midface tissues accentuates any weakness or laxity in the lower eyelid. This is important to recognize because a midface lift may be necessary to support the lower eyelid during blepharoplasty.

Jowl and Neck Region

After a thorough evaluation of the upper and middle third of the face, the surgeon should systematically evaluate the lower third of the face. Aging changes in the jowl and neck region include ptotic facial tissues at the mandibular border, bony resorption at the chin, and prominent submental fat. Atrophic redundant platysmal muscle may create vertically oriented paramedian bands in the neck. The location of the hyoid bone should be noted. Ideal patients have a superiorly positioned hyoid because a lower hyoid will limit the amount of refinement that can be achieved in the cervical mental angle. Ptotic submandibular glands are important to recognize preoperatively because they may appear more prominent after a rhytidectomy; this should be discussed with the patient. Significant actinic damage or fine wrinkling of the skin should be evaluated because these findings may require resurfacing procedures for improvement.

NONSURGICAL THERAPY

Some patients are well served by nonsurgical rejuvenation techniques. These offer less downtime, expense, and risk than more invasive surgical interventions. These modalities may benefit patients who have fewer advanced signs of aging and are good alternatives for patients who are not suitable surgical candidates. Furthermore, patients who are intimidated by the thought of surgery may be more comfortable with a nonsurgical treatment. Some of these procedures, however, require repeated administration because the benefits achieved lack longevity.

Prevention

Our society is recognizing the importance of health maintenance as we are living longer and healthier lives. Exercise, proper nutrition, and avoidance of detrimental extrinsic factors are beneficial to one's overall health, and these habits help prevent some of the changes associated with aging. Protecting skin from UV radiation in sunlight is essential to maintaining healthy skin. Avoidance and daily use of an appropriate sun block should be discussed with patients because these habits lessen the actinic changes associated with sun damage and decrease one's likelihood of developing cutaneous malignancies. Cosmetic products such as moisturizers that contain sun block are available and offer the benefits of systematic, regular use. Cigarette smoking is detrimental to the skin and body, and the benefits of smoking cessation should be discussed with patients. Finally, the benefits of routine skin care regimens should be discussed with patients, especially those seeking counsel earlier in life.

Chemodenervation

Active rhytids are caused by the action of underlying musculature. The rhytids are perpendicular to the axis of the muscle, just as relaxed skin tension lines are found perpendicular to underlying muscles. These rhytids are seen during muscle contraction and fade rapidly after the muscle relaxes. Typical patterns of dynamic facial rhytids are radial periorbital rhytids from the orbicularis oculi muscle (crow's-feet), horizontal forehead rhytids from the frontalis muscle, vertical glabellar rhytids from the action of the corrugator supracilii muscle (frown lines), and horizontal glabellar rhytids from the procerus muscle.

Botulinum neurotoxin is a natural toxin from *Clostridium botulinum* that causes muscle paralysis. The site of action is the presynaptic motor end plate (**Fig. 57–1A,B**). Binding of *C. botulinum* toxin prevents the release of acetylcholine. It takes 2 or 3 days for the effect to be observed, with peak paralysis occurring at around 1 week. Paralysis of the motor end plate is permanent, but resorption of blocked end plates and axonal sprouting cause the return of muscle function at ~4 to 6 months (**Fig. 57–1C**).

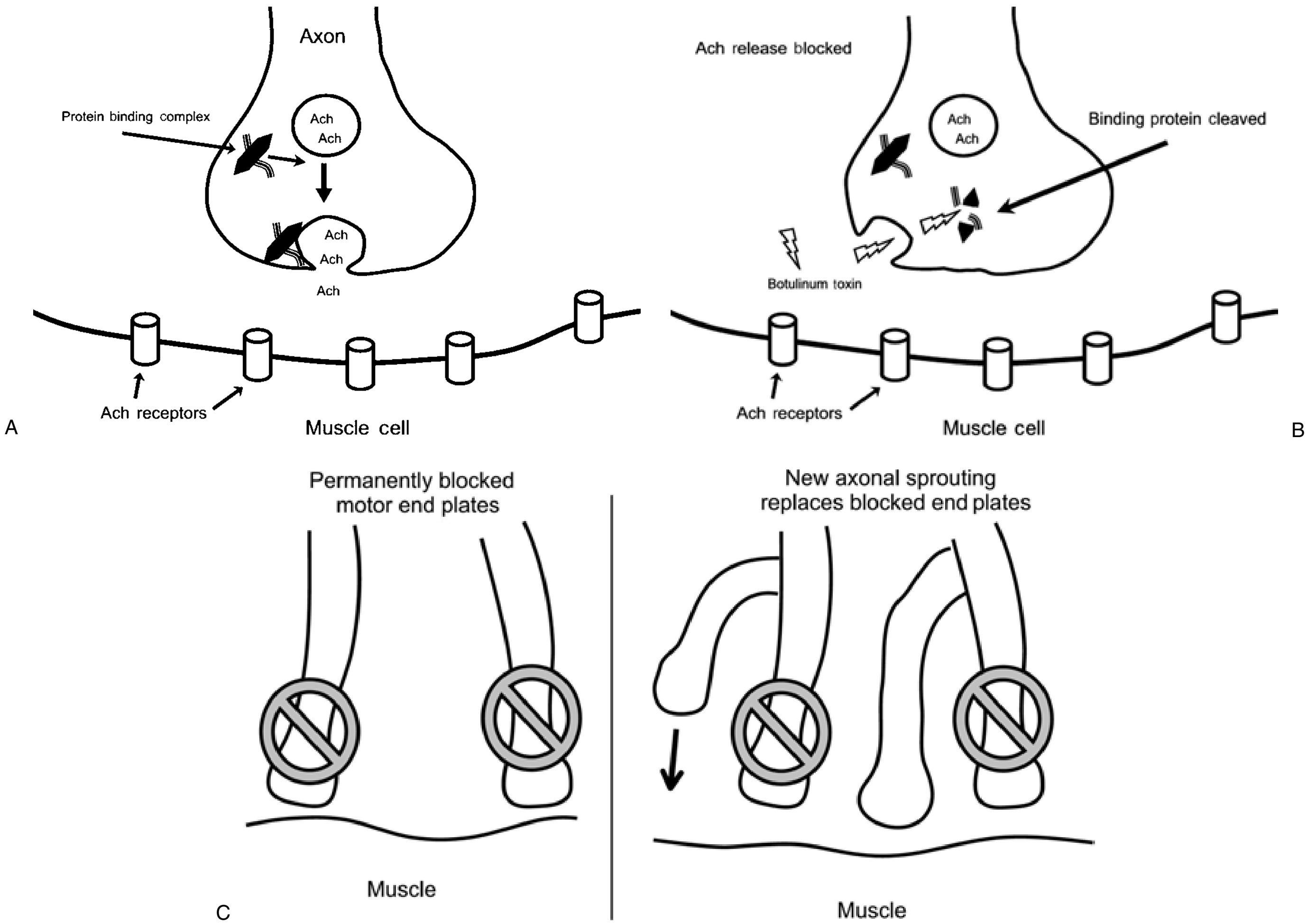


Figure 57–1 *Action of botulinum toxin at the motor end plate.* ***(A)*** *Normal motor end plate.* ***(B)*** *Blockage of release of acetylcholine vesicles by botulinum toxin.* ***(C)*** *Axonal sprouting with muscle reinnervation. ACh, acetylcholine.*

The most potent form of *C. botulinum* toxin is serologic type A. It is commercially available as Botox (Allergan, Inc., Irvine, CA), and has received approval from the U.S. Food and Drug Administration for use in facial rejuvenation. Each vial has 100 units, which is far below the median lethal dose of 2500 to 3000 units. The technique involves having the patient activate the target muscles, isolating the muscle between two fingers, then injecting the toxin into the muscle belly. Muscle paralysis results, with loss of dynamic rhytids in the affected area. Injection too close to the orbit may cause lid ptosis and potentially impaired ocular motility if the extraocular muscles are affected. Paralysis is self-limited, and injections are usually repeated every 4 to 6 months.

Resurfacing

Facial resurfacing is used to treat fine wrinkles and actinic changes associated with solar damage. The basic concept of most modalities is to ablate the epidermis and variable amounts of upper (papillary) dermis. The cells that have actinic damage are thus removed as well as portions of the damaged collagen in the supporting dermis. Regenerated dermis contains tight, supportive collagen bundles.

Preserving adenexal appendages is a key component to this process. Epithelial cells surrounding sebaceous glands, sweat glands, and hair follicles facilitate epithelial regeneration. The patient must be asked about systemic 13-*cis*-retinoic acid (Accutane) use or radiation exposure that can cause atrophy of the adenexal appendages, thereby increasing the risk of delayed healing and scarring. Most authors recommend stopping Accutane for at least 12 to 18 months prior to resurfacing procedures.

When using ablative techniques, the amount of improvement seen in resurfaced skin is proportional to the amount of ablation. Significant tightening results from ablation to the level of the reticular dermis, typically from multiple passes with a CO_2 laser or a Baker-Gordon peel. Unfortunately, with greater depths of injury, there are greater risks of complications. Disfiguring scarring can occur from overly aggressive ablation.

Postoperative care consists of open therapy with ointments or closed treatment with bio-occlusive dressings. Both aim to keep a moist, clean wound environment that allows rapid reepithelialization. In general, occlusive dressings provide more comfort but also increase the risk of infection if left in place for a prolonged period of time. Many surgeons use a combination of both types of treatments, placing a bio-occlusive dressing for the first 24 to 48 hours followed by open therapy for the next several weeks.

Pigmentary changes are expected with any procedure that penetrates the dermis. Phenol-containing peels are thought to be particularly toxic to melanocytes, increasing the risk of hypopigmentation. Furthermore, deeper ablations have a greater risk of hypopigmentation. Patients should be informed, however, that a lightening in skin color in the resurfaced region is expected. Consistency in technique is crucial because inconsistent depth of ablation can result in irregular pigmentation. In addition, patients seeking regional resurfacing should be counseled that the treated region may become noticeably lighter than the surrounding skin. Patient selection is important. Those with darker pigmentation, Fitzpatrick grade III or higher (see **Table 57–2**), will have a more dramatic demarcation between native and resurfaced skin. Furthermore, these patients are at greater risk for hyperpigmentation. Oral contraceptives and sun exposure shortly after resurfacing also increase the risk of hyperpigmentation. Bleaching creams may be used to treat hyperpigmentation with some success.

A herpes simplex infection can be induced whenever the perioral region is treated. Perioperative prophylaxis with acyclovir is recommended. Other common problems seen include prolonged erythema and milia formation. The most alarming complication is scarring, seen with excessive depth of ablation, infection, and delayed healing.

Chemical Peels

Chemicals agents are classified based on the depth of ablative injury. Superficial peels remove only the epidermis and have a relatively short recovery period. The amount of improvement seen is limited, as are the potential risks. Common agents include 15 to 20% trichloroacetic (TCA) acid and 70% glycolic acid. Superficial peels typically are performed without anesthesia in the office. For patients requiring slight to moderate improvements, repeated superficial peels can be performed on a scheduled basis to thin the stratum corneum and help maintain youthful, healthy skin.

Medium-depth peels ablate into the papillary dermis, inducing a greater increase in collagen density and a greater improvement in skin tightening. Healing typically occurs over several days. The most commonly used agent is 35% TCA. It is fairly safe and predictable, but at higher concentrations it can cause scarring. To improve the depth of peel, TCA can be combined with Jessner's solution or 70% glycolic acid as prepeel agents.

Deep peels have the greatest benefit in wrinkle improvement and skin tightening. The Baker-Gordon peel combines 3 mL 88% phenol, 2 mL H_2O, and 8 drops of septisol with 3 drops croton oil to cause deep penetration into the reticular dermis. Phenol at a concentration of 88% is a keratocoagulant. With dilution, it becomes keratolytic, allowing deeper penetration; hence the effectiveness of the Baker-Gordon formula. Phenol is toxic, and cardiac arrhythmias can be associated with resurfacing the face too rapidly. It is recommended not to resurface more than 50% of the face in a 60-minute period, allowing time for excretion and metabolism of the phenol. Hydration is used to decrease toxic reactions, and telemetry is recommended to monitor for cardiac arrhythmias. Hypopigmentation is common after Baker-Gordon peels, and patient selection is critical.

Laser Resurfacing

The CO_2 laser is the most common laser used for facial resurfacing. The wavelength of the coherent light is 10,600 nm. The depth of penetration is variable, depending on the machine, the settings, and the surgeon's technique. In addition to tissue ablation, there is thermal damage to the dermis immediately adjacent to the ablation zone. This damage is thought to enhance dermal tightening, but it can lead to complications. Prolonged erythema, sometimes lasting up to 3 months, is expected with ablation to the reticular dermis. Treatment with the erbium:YAG (yttrium-aluminum-garnet) laser offers ablation with less penetration and less collateral thermal damage. Recovery is quicker, with shorter periods of wound healing and erythema; however, the effects of isolated treatments are less dramatic than those achieved with the CO_2 laser.

Nonablative therapy is a recent adaptation of laser technology. Using a combination of a neodymium (Nd):YAG pulsed laser and cooling spray, the dermis is heated to 70°C, and the epidermal temperature is limited to 48°C. The result is preservation of the epidermis and thermal injury to the dermis that stimulates fibroblast activity and improvement in rhytids. The patient is spared the traditional extended recovery time associated with epidermal ablative lasers. In a clinical trial in which patients received three treatments to periorbital rhytids over a 3-month period, mild improvement was observed in 60% of patients. This modality appears safe even in Fitzpatrick skin types V and VI. Although the results have not been shown to be as dramatic as that achieved with ablative therapy, the limited morbidity, limited downtime, and good safety profile offer an appealing new option to people seeking improvement in facial rhytids.

The risks of laser resurfacing include all those found in chemical peeling. In addition, surgeons should attend a laser safety course prior to using lasers for skin resurfacing to learn about the management of laser fires as well as specific precautions necessary for the patient and the operating room staff.

Dermabrasion

Dermabrasion involves mechanical abrasion of the skin. Although it could be applied to most parts of the face, it is most often used to rejuvenate the perioral region, where it can be particularly effective at effacing rhytids. Typically, a diamond fraise wheel or wire brush on a hand engine is rotated at a speed appropriate to abrade the skin. It is important to keep the skin taut and rotate the bit toward the free border of the lip to avoid catching it. The perioral region is outlined and treated as a whole. Universal precautions should be exercised because there is potential for aerosolizing pathogens. Determining depth of ablation requires experience, and the appearance of the reticular dermis is described as being like a chamois cloth. The standard viral prophylaxis and postoperative care with open or closed therapy are recommended. Complications from dermabrasion are similar to those seen with other resurfacing techniques.

Injectable Soft Tissue Fillers

An effective method to treat contour deformities such as deep static wrinkles, prominent nasolabial folds, atrophic lips, and depressed scars is to use injectable substances for augmentation. Currently, there are numerous substances marketed as injectable fillers, including autologous fat, autologous collagen, hyaluronic acid, and polymethylmethacrylate microspheres. Two commonly used materials are bovine collagen and particulized cadaveric acellular dermis. Because the bovine collagen is a xenograft, an allergic reaction is possible, and skin testing is required. Acellular dermis is similar in qualities to collagen without the concerns for an allergic response. Both products are absorbed over time and require repeat injections to maintain benefit.

SURGICAL TREATMENT

Surgical intervention is often required for significant repositioning and recontouring of facial tissues. Aging face surgery is often performed on an outpatient basis using deep sedation or general anesthesia.

Browlift

There are multiple surgical approaches used to lift the ptotic brow. Factors that influence the type of browlift employed include the location of the hairline, the presence of alopecia or hair thinning, the amount of lift required, and the shape of the forehead. The location of the hairline is important because both coronal lifts and endoscopic lifts will tend to raise the hairline. Thus a high hairline is a relative contraindication to these procedures, and a trichophytic, midforehead, or direct lift should be considered. Thinning hair or male pattern baldness creates a problem with scar camouflage. In these cases, the short incisions of an endoscopic lift or a midforehead lift with scars hidden in horizontal forehead creases would be preferred.

With coronal and endoscopic browlifts, the plane of dissection is subperiosteal or subgaleal. The supratrochlear and supraorbital neurovascular bundles are carefully identified and preserved to prevent forehead and scalp dysesthesias. The ligamentous attachments at the orbital rim should be elevated carefully to allow sufficient elevation. Dissection laterally extends to the zygomatic arch with blunt spreading to avoid injury to the frontal branch of the facial nerve. Corrugator muscles may be carefully excised to address glabellar rhytids.

Coronal lifts require an incision from ear to ear across the scalp; as a result, postoperative vertex hypoesthesia is a common complaint. Skin is excised with coronal lifts, enhancing the effectiveness and durability of the lift. Endoscopic forehead lifting avoids a large incision by using multiple small ports to facilitate endoscopic dissection. Suspension is achieved by various methods, including screws, fibrin glue, and sutures. Some surgeons question the long-term durability of endoscopic lifts. Nonetheless, endoscopic browlifts have become very popular, with good patient and surgeon satisfaction. Patients should be informed that the expected amount of lift with an endoscopic approach is ~5 to 10 mm. Patient selection is important, with the best candidates for endoscopic browlifts being those with low or normal hairlines, thinner skin, and no significant forehead contour abnormalities.

Patients who are not suitable candidates for an endoscopic forehead lift may benefit from a trichophytic lift.

The trichophytic incision is made just anterior to the anterior hairline. Care is taken to bevel the blade to allow hair regrowth through the scar. The trichophytic incision can be combined with subperiosteal dissection through a midline incision in the frontalis muscle. This dissection facilitates release of supraorbital ligamentous attachments, while avoiding damage to the deep branch of the supraorbital nerve. This branch runs in the subgaleal plane in a lateral position and is severed during a coronal lift, leading to persistent vertex scalp dysesthesias.

The midforehead lift uses horizontal rhytids to camouflage the scar. Direct browlifts place the incision directly over the brows, allowing precise control when treating asymmetric brows. The midforehead and direct browlifts have a greater potential for more noticeable scars, but they are beneficial in addressing brow ptosis associated with facial paralysis.

Blepharoplasty

Upper lid blepharoplasty involves excision of redundant upper eyelid skin, orbicularis oculi muscle, and/or preaponeurotic fat. Fat can be removed from the medial and central orbital fat pads if fat pseudoherniation is present. Laterally, the lacrimal gland may become ptotic and require resuspension. Meticulous hemostasis is achieved prior to closure to avoid potentially devastating postoperative complications associated with hematomas.

The lower lid can be approached through a transconjunctival incision, providing access to the lower orbital fat compartments (i.e., medial, middle, and lateral). The anterior lamella is preserved with this technique, thereby decreasing the likelihood of lid retraction secondary to scarring of eyelid skin and muscle. An important structural component of the lower eyelid, the orbital septum, represents a continuation of the periosteum of the orbital floor and maxilla extending up between the orbicularis muscle and conjunctiva of the lower lid. The orbital septum weakens with aging, allowing pseudoherniation of orbital fat. Some patients may benefit from "tightening" of the orbital septum through application of bipolar cauterization. Fat also can be removed from the lower orbital fat compartments if more significant fat protrusion is noted. If lower eyelid skin needs to be excised, it is typically done with a "pinch" technique when transconjunctival blepharoplasty is performed.

Another approach to lower lid blepharoplasty involves using a subciliary skin incision. Orbital fat is managed in a manner similar to that described with a transconjunctival approach. Redundant skin and muscle can be excised as the skin flap is elevated and advanced superolaterally. If significant dermatochalasis or hypertrophied orbicularis oculi muscle is noted on preoperative examination, this approach is often used to maximize tightening of the lower eyelid skin and muscle.

Complications associated with upper lid blepharoplasty include hematoma, lagophthalmos, and levator dysfunction. Upper eyelid skin should be stored after excision, in the event that excessive skin was removed and grafting is required. Most cases of lagophthalmos respond to conservative treatment, including lid massage and eye care. Levator injury may improve with conservative treatment, but revision surgery with levator repair may be necessary.

Hematomas are a concerning occurrence. If limited to the lid, these can be drained by opening the incisions and treating the bleeding site. If the bleeding dissects posteriorly, a retrobulbar hematoma can result and requires immediate attention. Assessment of visual acuity is performed. If there is visual impairment and the globe is proptotic and firm, a canthotomy and cantholysis should be performed. During the immediate postoperative period, sutures should be removed from surgical eyelid wounds, and any clot evacuated. Medical therapy with steroids, diuretics, and mannitol is initiated while an ophthalmology consultation is requested. Prevention is critical. Avoiding medications that prolong bleeding times and controlling blood pressure perioperatively are extremely important. Meticulous hemostasis is also essential. Blindness occurs with an incidence of 0.04%. It can be related to development of a retrobulbar hematoma, excessive retraction on the globe, or vasospasm from epinephrine in the local anesthetic.

The most common complication of lower lid blepharoplasties is lid malposition (retraction or ectropion). The best treatment is prevention with careful preoperative assessment and appropriate lid-tightening procedures at the time of the blepharoplasty. If lid malposition is noted postoperatively, massage of the lower eyelid may restore eyelid position. Patients should be started on a regimen of eye care, including lubricating drops. Skin grafting and lid tightening procedures may be necessary in severe cases of eyelid malposition.

Extraocular motility can be affected by lower lid blepharoplasty. The inferior oblique muscle separates the medial and central fat compartments and can be damaged during fat removal or cauterization.

Midface Lift

A variety of procedures have been developed to address ptosis of the malar and suborbicularis oculi fat pads.

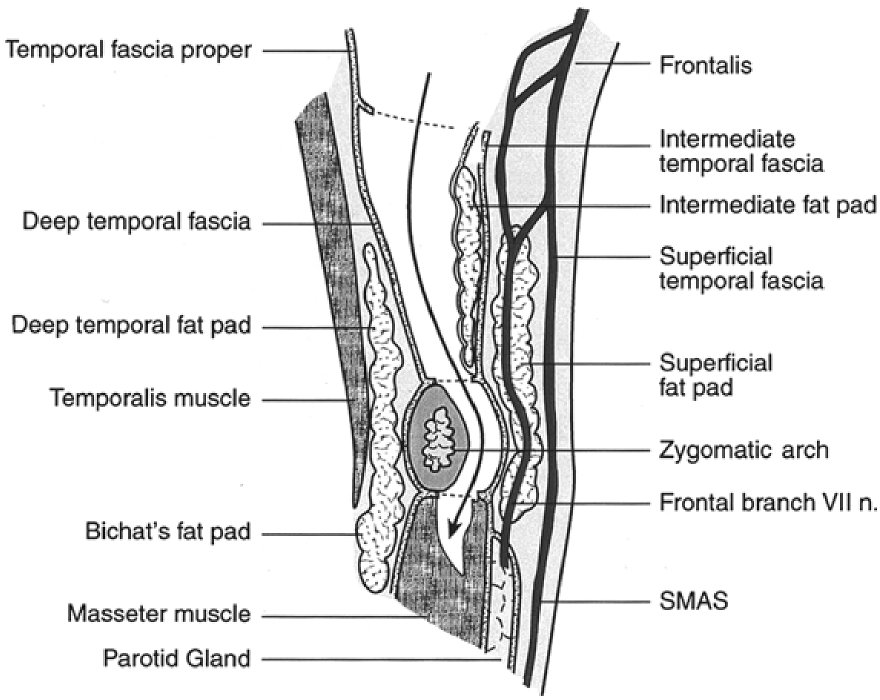


Figure 57–2 *Relationship of fascial planes, frontal branch of the facial nerve, and temporal fat pads at the level of the zygomatic arch. SMAS, superficial musculoaponeurotic system. (From Papel ID, Frodel J, Holt GR, et al. Facial Plastic and Reconstructive Surgery. New York: Thieme Medical Publishers; 2002.)*

These procedures have a common purpose—to elevate ptotic soft tissues in the infraorbital and cheek region. Midface lifts are often combined with lower eyelid blepharoplasty, browlift, and rhytidectomy.

Soft tissue overlying the maxilla and medial zygoma is dissected, usually in a subperiosteal plane. Some techniques isolate and elevate malar and/or suborbicularis oculi fat pads, without performing a subperiosteal dissection. The midface can be approached endoscopically through a temporal incision, and this approach is often combined with an endoscopic brow lift. The frontal branch of the facial nerve can be injured during dissection in the temporal area (**Fig. 57–2**). Release of ligaments at the lateral orbital rim is important to maximize potential elevation. Midface tissues are resuspended in a vertical direction. Postoperative edema of the midface tissues is expected to persist for ~4 to 6 weeks. Complications include injury to the infraorbital nerve, scleral show, and lower lid malposition.

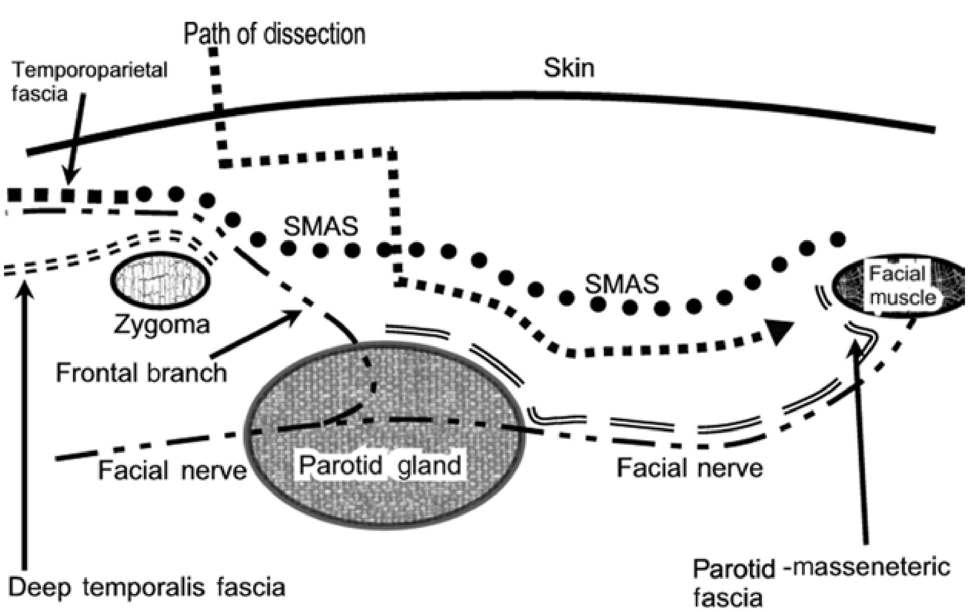


Figure 57–3 *Level of dissection for sub-superficial musculoaponeurotic system (SMAS) facelift.*

RHYTIDECTOMY

Traditional facelift techniques elevated subcutaneous flaps; however, benefits were short lived, as the skin stretched and scars widened over time. The SMAS lift was developed after identifying the importance of the SMAS in facial movement and soft tissue support. The level of dissection is below the SMAS, but above the parotidomassenteric fascia and facial nerve (**Fig. 57–3**). The SMAS is resuspended in a vertical and lateral direction. There are many variations of the SMAS lift, but its major advantage is applying tension to the SMAS rather than the overlying skin. This provides more durability, recontouring of subcutaneous tissues, and less tension on the skin closure. One disadvantage of the traditional SMAS lift is that surgery only addresses the jawline and anterior neck. The midface is not addressed, leaving some patients with prominent nasolabial folds and ptotic midface soft tissue. Some surgeons perform midface lifts in conjunction with SMAS lifts to address this concern. In addition, the deep plane and composite facelifts are two procedures that combine sub-SMAS dissection with techniques designed to correct midface ptosis. The deep plane lift extends dissection over the zygomaticus major muscle, releasing attachments of the SMAS to the overlying nasolabial fold. This facilitates effacement of the nasolabial fold; furthermore, proponents emphasize improved flap vascularity because the entire facial flap is raised in a sub-SMAS fashion. The composite facelift is similar to the deep plane facelift; however, dissection in the periocular region is performed in a submuscular plane to facilitate repositioning of the orbicularis oculi muscle.

Facelifts are often combined with neck dissection to address prominent submental fat, platysmal banding, and significant jowling. Platysmal plication and removal of submental fat are performed through a submental incision. Other procedures that can be combined with facelifts to optimize aesthetic outcomes include placement of prejowl and chin implants to improve skeletal contour.

The most common complication of facelifts is hematoma, with an incidence of 3.6%. Patients complain of worsening pain, swelling, and ecchymosis. Hematomas should be promptly drained and treated. The most common nerve injured is the greater auricular nerve, and careful, superficial dissection over the sternocleidomastoid muscle will reduce this risk. The most common motor nerves injured are the frontal and marginal mandibular branches of the facial nerve. Skin flap necrosis can occur in the posterior auricular area, and smokers are at significantly higher risk for this problem. A pixie ear deformity results from poor incision planning and excessive tension on the earlobe from the cervical skin flap.

SUMMARY

As society becomes more knowledgeable about treatment of the aging face, it is important for physicians who treat these patients to maintain a high standard of competency. This requires a thorough understanding of facial anatomy and physiology, which is the basis for therapeutic intervention and benefit.

SUGGESTED READINGS

Hamra ST. Composite Rhytidectomy. St. Louis: Quality Medical Publishing; 1993

Johnson CM, Alsarraf R. The Aging Face: A Systematic Approach. Philadelphia: WB Saunders; 2002

Papel ID, Frodel J, Holt GR, et al. Facial Plastic and Reconstructive Surgery. New York: Thieme Medical Publishers; 2002

Romo T, Millman AL. Aesthetic Facial Plastic Surgery. New York: Thieme Medical Publishers; 2000

SELF-TEST QUESTIONS

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. When performing a cervical rhytidectomy, the surgeon should recognize that the plane of dissection is
   A. Superficial to the facial nerve at all points
   B. Deep to the facial nerve, inferior to the mandible
   C. Between the subdermal plexus and dermis
   D. Between the platysma muscle and marginal mandibular nerve

2. Twenty minutes after arrival in the recovery room, a patient notes double vision and severe unilateral left eye pain following four-quadrant blepharoplasty. On exam the left eye is fixed, proptotic, with periorbital ecchymosis and decreased acuity when compared with the preoperative exam. What is the best course of action?
   A. Apply ice to involved eye and allow local anesthetic to wear off.
   B. Perform an immediate canthotomy/cantholysis, open incision lines, administer mannitol and a diuretic, and request an urgent ophthalmology consult.

C. Open incision lines to evacuate clot and ask for operating room time for wound exploration.
D. Administer mannitol, decadron, and a diuretic while awaiting an ophthalmology consult.

3. A 37-year-old patient presents with complaints of vertical rhytids in the glabellar region as well as horizontal rhytids on the forehead. Her brow position is appropriate for a female without ptosis noted, though her hairline is high. Her skin is a Glogau class II, with wrinkles noted on expression. She is treated with botulinum toxin injections and is happy with her results. She returns 6 months later, and her dynamic rhytids have recently returned. How did this occur?

A. Acetycholine overproduction has compensated for blockage of acetycholine release at presynaptic motor end plate.
B. Muscle cell regeneration has occurred.
C. Resorption of blocked end plates and axonal sprouting with reinnervation have caused a return of muscle function.
D. An improper dose of botulinum toxin was administered, limiting the longevity of muscle paralysis.

Chapter 58

VASCULAR ANATOMY OF THE HEAD AND NECK

Jane A. Petro

GENERAL PRINCIPLES OF HEAD AND NECK CIRCULATION

The adult anatomy of blood vessels of the head and neck can best and most simply be visualized as a series of concentric circles. The extracranial and intracranial vessels share a rich network of connections that loop from side to side and in and out of the skull. Few of the vessels are true terminal arteries, and young, healthy individuals will rarely exhibit any effect from the ligation of any single major trunk. Branch ligation will rarely cause ischemia, even in those with advanced arteriosclerosis. The common carotid, vertebral, and subclavian arteries provide the blood supply to the head and neck. Their multiple interconnected branches include (1) the inferior thyroidal vessels from the subclavian linking to the superior thyroidal branch of the external carotid; (2) the internal carotid vessels to the vertebral arteries via the circle of Willis; and (3) the internal carotids via the ophthalmic artery branch from the circle of Willis to the orbital branches of the facial and temporal branches of the external carotid. The accompanying veins share the names of their arterial counterparts and provide the primary drainage into the internal jugular system. The external jugular vein is unique, lacking an accompanying artery.

DISTINCTIVE PATTERNS OF THE BLOOD SUPPLY (FIGS. 58–1 AND 58–2)

The circulation of the skin of the head and neck illustrates the distinctive characteristics of the blood supply necessary for flap design (see Chapter 59), defined as axial, musculocutaneous, or random pattern. Vascular territories may accompany named vessels, such as the temporal artery, which are called axial flaps, or they may result from musculocutaneous perforators, as in the platysma. Random pattern flaps are based on the dermal and subdermal plexus of vessels. The use of such flaps in head and neck surgery depends on knowledge of anatomical territories appropriate to aesthetic considerations as well as to vascular patterns. The original axial flap, the Bakamjian deltopectoral flap, was based on the intercostal perforators of the internal mammary artery. This was the original axial flap used in reconstruction of the head and neck.

A more profound understanding of the skin's blood flow has been elucidated over the past 30 years as interest in flap anatomy, design, and free flap execution has created a clinical relevance not previously necessary (see Chapter 59). Taylor and his coworkers (1994) proposed the concept of the angiosome, defining the arterial system of the skin and underlying tissues. This concept recognizes that the arterial supply of the skin follows the cutaneous innervation. The work done by Taylor et al detailed in their human and comparative animal studies the finding that cutaneous arteries may sometimes accompany nerves, but nerves are always accompanied by an artery. The design of flaps, then, may be dependent on their neurocutaneous supply if innervation is a consideration. Such flaps may be more relevant to reconstructive surgery of the extremities than of the head and neck. However, the principle that nerves and vessels accompany each other is well illustrated in the head, with named arteries accompanying the sensory branches of the seventh cranial nerve (CN VII; e.g., alveolar, supraorbital, infraorbital, supratrochlear).

THE ADULT PATTERN OF THE ARTERIAL VASCULAR TREE (FIGS. 58–3 AND 58–4)

The aortic arch gives rise directly to the left common carotid and the brachiocephalic artery from which the right common carotid branches. The left common carotid, in one variation described as bovine, may arise from the innominate brachiocephalic artery. There are no branches of the common carotid until it bifurcates into the internal and external carotids. There are no branches of the internal carotid until it passes through the carotid canal into the skull.

The vertebral artery arises as the first branch of the subclavian artery and goes posterior and medial to enter the foramen transversarium of the sixth cervical vertebra. The vertebral artery has two segments, the cervical portion with branches to the spinal musculature, and the cranial portion with branches to the meningeal, posterior spinal, and anterior spinal arteries. The right and left vessels pass along the vertebra through the foramen magnum, joining together to form the basilar artery.

Branches of the vertebral artery include

- Anterior spinal artery
- Posteroinferior cerebellar artery
- Basilar artery
- Anteroinferior cerebellar artery
- Pontine perforating branches
- Posterior cerebellar artery

The basilar artery then rebifurcates to form the right and left posterior cerebral arteries, which form part of the circle of Willis. Additional intracranial blood supply comes from a branch of the external carotid, the maxillary artery that provides the middle meningeal artery to the dura. The circle of Willis lies between the cerebral hemispheres and the optic chiasm. Posteriorly, it is composed of the bifurcation of the basilar artery forming the P1 segment of the posterior communicating artery, linking the middle cerebral artery, then anterior to the horizontal A1 anterior communicating artery, which gives off the ophthalmic artery passing forward, and joining across the midline as the anterior communicating artery.

The lower neck is supplied by the thyrocervical arterial trunk, a branch derived directly from the subclavian artery. This branches in turn to form the inferior thyroidal artery and the suprascapular artery. There are numerous variations in the branching of this arterial system.

The remaining neck, face, scalp, tongue, maxilla, and mandible are supplied by the external carotid artery, whose chief branches in order (most common finding, with variations being the rule rather than the exception) are (1) superior thyroidal artery, which supplies the superior thyroid and the superior laryngeal vessel; (2) ascending pharyngeal artery, which supplies the pharyngeal wall and palate with branches that include the meninges through the jugular foramen, foramen lacerum, and the hypoglossal canal (interconnections between the ascending pharyngeal artery, the occipital artery and the vertebral artery are common); (3) lingual

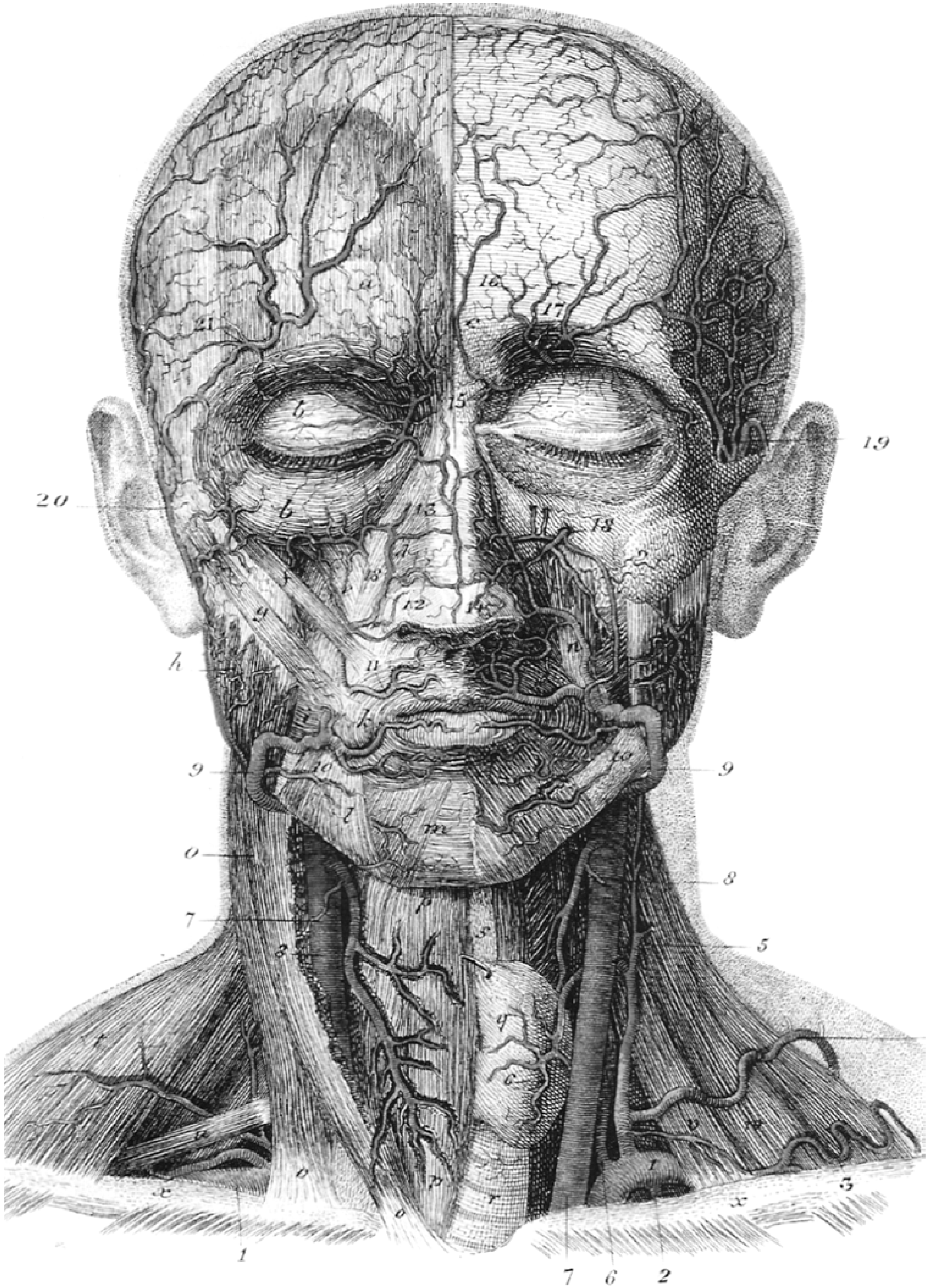


Figure 58–1 *Muscles and arteries of the anterior surfaces of the head and neck. Muscles: (a) occipitofrontalis; (b) orbicularis palpebrarum; (c) corrugator supercillii; (d) levator labii superioris alaeque nasi; (e) levator labii superioris proprius; (f) xygomaticus minor; (g) xygomaticus major; (h) masseter; (i) buccinator; (k) obicularis oris; (l) triangularis menti; (m) quadratus menti; (n) levator anguli oris; (o) sternocleidomastoideus; (p) sternohyoideus; (q) glandula thyroidea; (r) trachea; (s) larynx; (t) cucullaris v. trapezius; (u) omohyoideus; (v) scalenus anticus; (w) scalenus medius; (x) clavicula. Arteries: (1) subclavia; (2) mammaria interna; (3) transversa scapulae; (4) transversa colli; (5) cervicalis ascendens; (6) thyroidea inferior; (7) carotis communis; (8) thyroidea superior; (9) maxillaris externa v. labialia; (10) coronaria labii inferioria; (11) coronaria labii superioria; (12) angularia; (13) dormalea nasi; (14) alares nasi; (15) opthalmica (with palpebralea); (16) frontalis; (17) supraorbitalis; (18) infraorbitalis; (19) temporales profundo (from the maxillaris interna); (20) temporalis (superficialis); (21) ramus frontalis a. temporalis.*

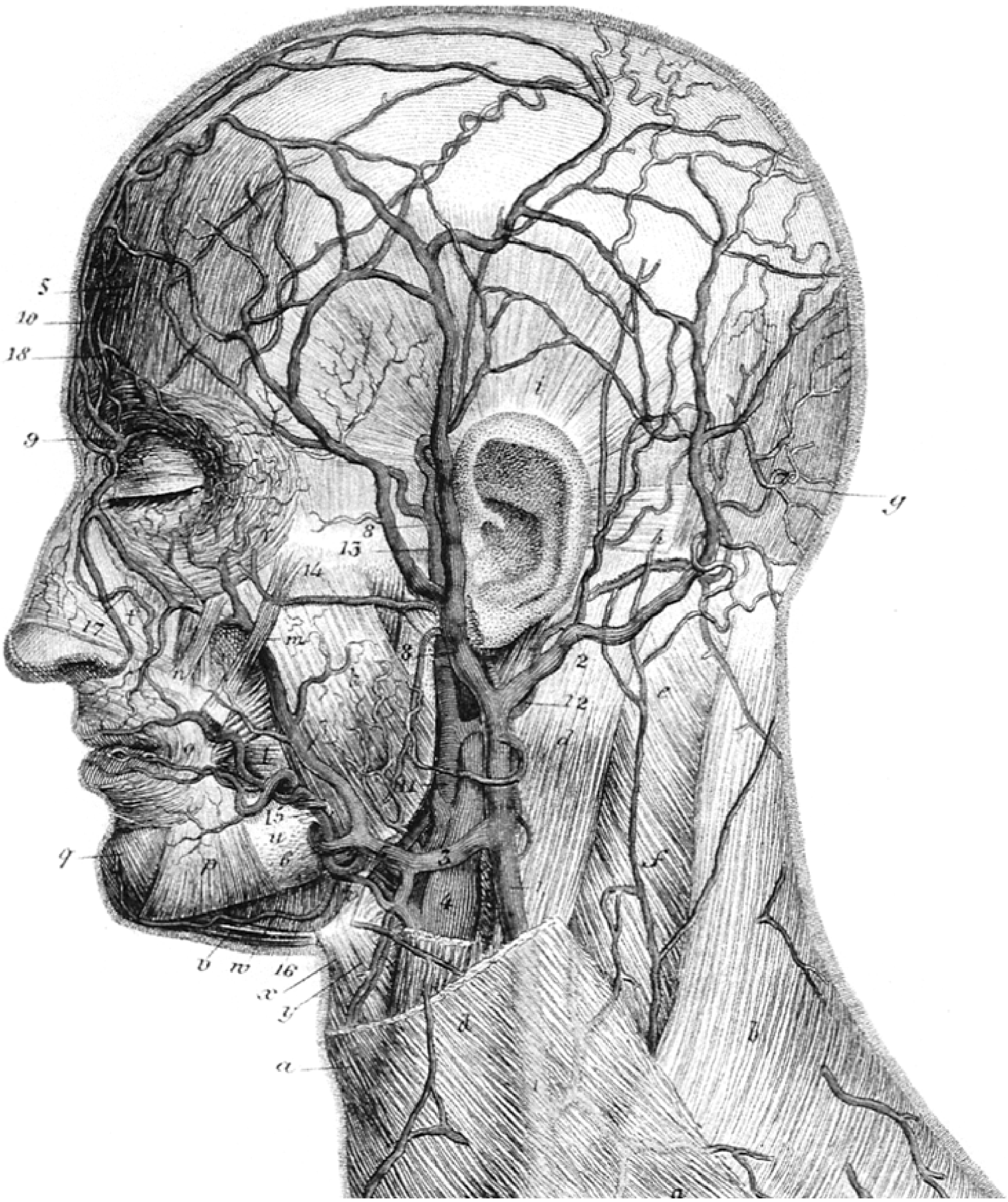


Figure 58–2 *Muscles, veins, and arteries of the lateral surface of the head, face, and neck. Muscles: (a) platysmamyoides v. latissimus colli; (b) cucullaris v. trapezius; (c) deltoidens; (d) sternocleidomastoideus; (e) splenius capitis; (f) splenius colli; (g) occipitalis; (h) retrahentes auris; (i) attollens auris; (k) masseter; (l) buccinator; (m) zygomaticus major; (n) zygomaticus minor; (o) orbicularis oris; (p) triangularis menti; (q) quadratus menti; (r) orbicularis palpebrarum; (s) frontalis; (t) levator labii superioris alaeque nasi; (u) maxilla inferior; (v) digastricus maxillae inferioris; (w) mylohyoideus; (x) sternohyoideus; (y) omohydoideus. Veins: (1) jugularis externa; (2) occipitalis; (3) ramus communicans inter v. jugularis externa et interna; (4) jugularis interna; (5) facialis interna; (6) labialis; (7) angularis; (8) temporalis; (9) ophthalmica cerebralis; (10) frontalis. Arteries: (11) carotis externa; (12) auricularis posterior; (13) temporalis (superficialis); (14) transversa faciei; (15) maxillaris externa, labialis v. facialis; (16) submentalis; (17) angularis; (18) frontalis.*

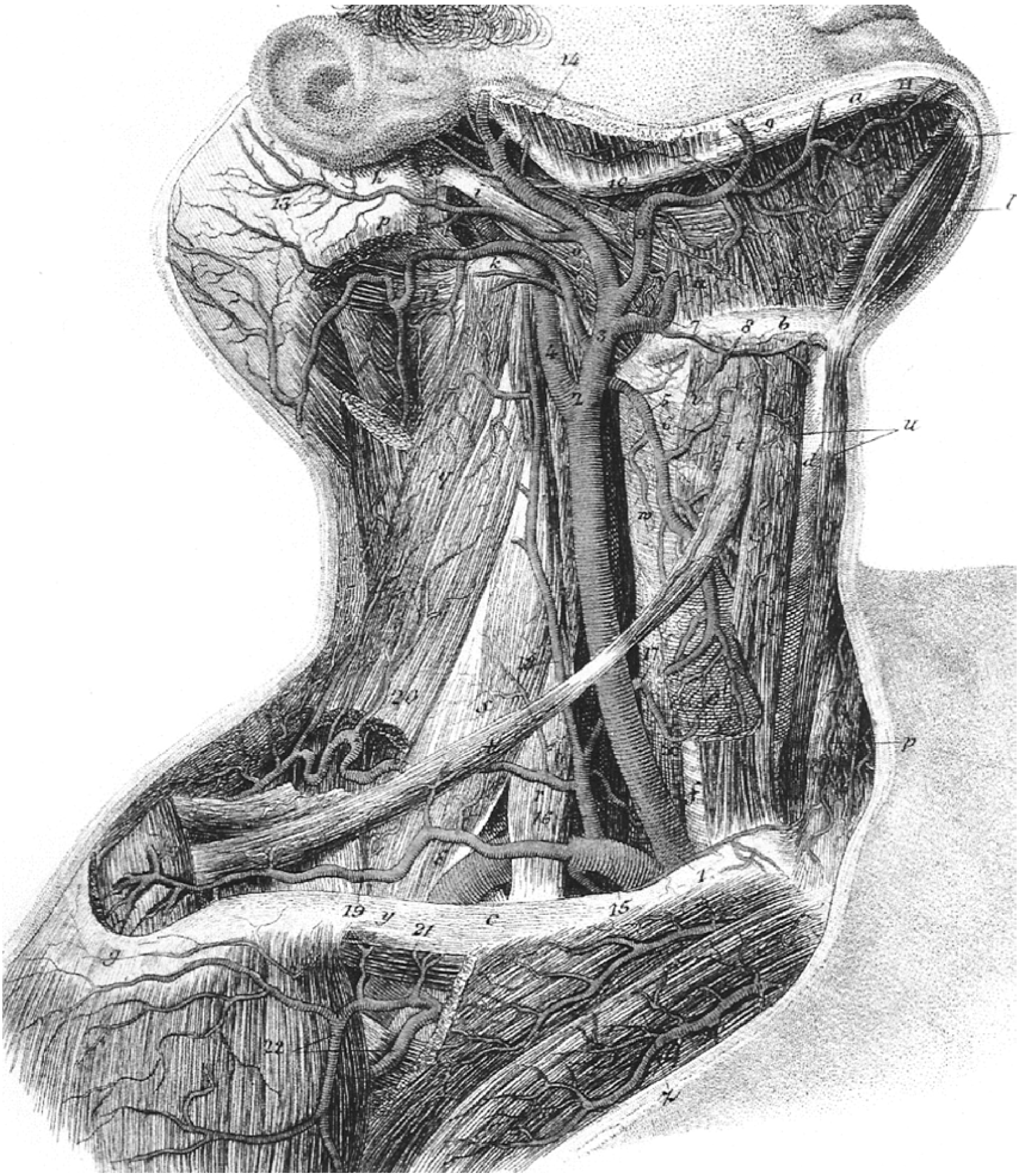


Figure 58–3 *Structures and arteries of the right side of the neck. Structures: (a) maxilla inferior; (b) os hyoides; (c) clavicula; (d) larynx; (e) glandula thyroidea; (f) trachea; (g) acromion scapulae; (h) processus mastoideus; (i) processus styloideus; (k) processus transversus atlantis. Muscles: (l) digastricus (venter anterior); (m) mylohyoideus; (n) hyoglossus; (o) styloglossus; (p) sternocleidomastoideus; (q) levator anguli scapulae; (r) scalenus anticus; (s) scalenus medius; (t) omohyoideus; (u) sternohyoideus; (v) thyrohyoideus; (w) pharynx; (x) esophagus; (y) subclavius; (z) pectoralis major. Arteries: (1) carotis communis dextra; (2) bifurcation of a. carotis communis dextra; (3) carotis externa; (4) carotis interna; (7) lingualis; (8) ramus hyoideus a. lingualis; (9) maxillaris externa v. facialis; (10) palatina ascendens; (11) submentalis; (12) occipitalis (with descending branches); (13) auris posterior; (14) temporalis (superficialis); (15) subclavia dextra; (16) truncus thyrocervicalis; (17) thyroidea inferior; (18) cervicalis ascendens; (19) transversalis humeri; (20) transversalis colli; (21) axillaris; (22) thoracicae externa.*

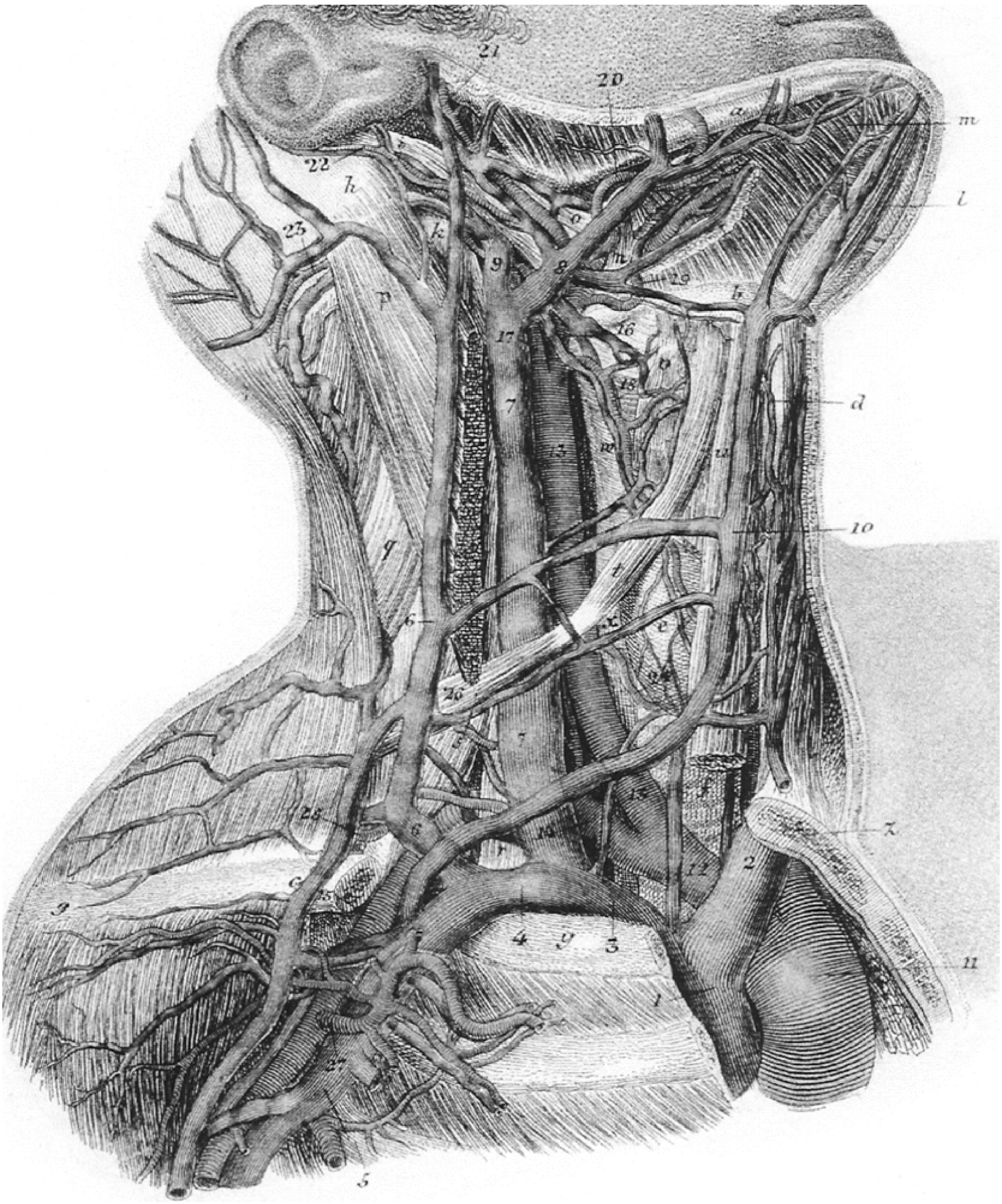


Figure 58–4 *Structures, muscles, veins, and arteries of the right side of the neck. Structures: (a) maxilla inferior; (b) os hyoides; (c) clavicula; (d) larynx; (e) glandula thyroidea; (f) trachea; (g) acromion scapulae; (h) processus mastoideus; (i) processus styloideus; (k) processus transversus atlantis. Muscles: (l) digastricus (venter anterior); (m) mylohyoideus; (n) hyoglossus; (o) styloglossus; (p) sternocleidomastoideus; (q) levator anguli scapulae; (r) scalenus anticus; (s) scalenus medius; (t) omohyoideus; (u) sternohyoideus; (v) thyrohyoideus. Structures: (w) pharynx; (x) esophagus; (y) first rib, first bone of the sternum. Veins: (1) cava superior; (2) innominata sinistra; (3) innominata dextra; (4) subclavia dextra; (5) axillaria; (6) jugularis externa; (7) jugularis interna; (8) facialis; (9) maxillaris interna; (10) jugularis media. Arteries: (11) arcus aortae; (12) innominata; (13) carotis communis dextra; (14) subclavia dextra; (16) carotis externa; (17) carotis interna; (18) thyroidea superior; (19) lingualia; (20) maxillaris externa v. facialis; (21) temporalis; (22) auricularis posterior; (23) occipitalis; (24) thyroidea inferior; (25) transversalis humeri; (26) transversalis colli; (27) thoracicae externa.*

artery, which supplies the hyoid bone and musculature, hypoglossus muscle, tongue, sublingual gland, and pharynx; (4) facial artery, which supplies the submandibular gland, the superior constrictor, and the stylohyoid and digastric muscles, crosses the anterior border of the mandible anterior to the parotid and deep to the mandibular branch of the facial nerve, and branches into the submental artery, the labial artery, and its terminal branch the angular artery, paralleling the nasolabial fold and passing along the nose toward the orbit; (5) occipital artery, which arises opposite (on the posterior wall of the external carotid) the facial artery and passes below as it supplies the sternocleidomastoid muscle into the posterior scalp, supplying the occipital muscles, posterior skin, and meninges; (6) posterior auricular artery, which branches first as the stylomastoid artery, passing through the stylomastoid foramen to supply the tympanic cavity, mastoid, and semicircular canals (the main vessel continues on to supply the posterior auricular skin and retroauricular scalp); (7) superficial temporal artery, which is the terminal branch of the external carotid artery, dividing from the maxillary artery within the parotid gland and supplying the anterior two thirds of the scalp, pericranium, and the temporalis muscle and fascia, portions of the parotid gland, anterior ear, and the temporomandibular joint; (8) maxillary artery, which has three main segments as it passes deep to the mandible and the infratemporal fossa through the pterygoid muscle. The first segment includes the inferior alveolar artery, then the middle accessory meningeal arteries, which pass through the foramen spinosum and ovale, the deep auricular branch to the external auditory canal, and the anterior tympanic artery. The second segment supplies the pterygoid and temporalis muscles and the lingual and buccal nerves. The third segment passes to the pterygopalatine ganglion and terminates in multiple branches to the intranasal cavity, the posterior superior alveolar artery to the palate and posterior wall of the maxilla, and finally the infraorbital artery through the anterior maxillary wall via the infraorbital foramin, where it may communicate with the facial artery.

The lower face is supplied by the facial artery, rising from the anterior aspect of the external carotid, between 1 and 3 cm above the common carotid bifurcation and within 1 cm of the lingual artery. Branches of the facial artery supply the neck, including the ascending palatine artery, the submandibular gland, variable branches to the platysma, and the submental artery. This branch anastomoses about half of the time with the sublingual artery, just beyond its perforation of the mylohyoid muscle. The vessel is quite variable as it enters the face, occasionally terminating just beyond the turn it makes around the mandible, or more commonly running toward the midface, to the nose. It carries inferior and anterior branches to the masseter muscle, as well as small muscular branches in the cheek to the various muscles of facial expression. At the nasolabial fold it branches into the inferior, then superior coronary vessels, which in turn supply the labial arteries. These run horizontally along the inner mucosa of the lip, with vertical branches to the orbicularis, and communications superiorly to the columella, septum, and ala. Beyond the coronary it continues as the angular artery (considered the terminal branch of the facial artery), supplying the nose, and communicating commonly with the internal nasal artery or ophthalmic artery, or both, or neither.

THE ADULT PATTERN OF THE VENOUS ANATOMY OF THE HEAD AND NECK (FIG. 58–4)

The venous drainage of the intracerebral circulation begins as the brain's venous drainage forms the superior and inferior sagittal sinuses. These communicate with the transverse and sigmoid sinuses to form the internal jugular vein. The ophthalmic veins drain both intracranially to the cavernous sinus and extracranially to the facial vein. Emissary veins connect the dural venous sinuses and the scalp veins. The cavernous sinus drains into the pterygoid plexus. The superior petrosal sinus and transverse sinus drain to the internal jugular vein that passes into the carotid sheath and through the neck to join the subclavian vein. The subclavian veins form the right and left brachiocephalic veins, respectively. The maxillary and superficial temporal veins join to form the retromandibular vein that communicates with both the internal and the external jugular vein. The superior and inferior thyroidal veins enter the internal jugular vein. The external jugular vein remains subcutaneous to the sternocleidomastoid muscle and joins the subclavian vein at the base of the neck. The anterior jugular vein drains the submental area and descends paralleling the sternohyoid muscle. The anterior jugular, external jugular, and internal jugular veins communicate through the communicating vein, which runs along the anterior border of the sternocleidomastoid muscle. Most of the named arteries of the head and neck have accompanying veins bearing the same name.

THE REGIONAL VASCULAR ANATOMY AFFECTING FLAP DESIGN AND SURGICAL CONSIDERATIONS

The facial artery and its multiple branches, with direct communications across the midline and via the infraorbital artery and mandibular alveolar artery, provide a rich arterial arcade. The common incisions used in facial surgery, parotidectomy, facelift, local rotation flaps used to close defects, and so on, are reliable as random pattern flaps. Tissue loss secondary to the complications of surgery in the area are caused by factors such as smoking, radiation, hematoma, and significant infections (i.e., those external factors that can independently affect blood flow).

Lips

The labial arterial vessels run along the intraoral mucosal surface of the lip, paralleling the vermilion border of the external mucocutaneous junction. The Abbe and Eastlander flaps are based on the labial artery, and although the presence of the vessel is variable, the rich interconnections in the mucosa and orbicularis muscle make these reliable.

External Ear

Blood supply to the ear comes from the external carotid artery, posteriorly via a direct branch called the posterior auricular artery, and anteriorly, from direct auricular branches of the superficial temporal artery. Additional supply comes to the posterior, upper portion of the ear from branches of the occipital artery that may also communicate with the posterior auricular artery. Success of replantation of the external ear or its reconstruction relies on identifiable, available branches of these arteries. Flaps designed in this area may be of an axial or a random pattern.

Periocular

The adnexa of the eye is supplied laterally and superiorly by branches of the superficial temporal artery, inferiorly by the transverse facial artery (a branch of the superficial temporal artery), and medially by the nasal artery. Just as the labial arteries of the lips parallel the vermilion margin, the superior and inferior palpebral arteries form a circle along the margin of the upper and lower eyelids. The rest of the blood supply to the area comes from branches of the ophthalmic artery, a branch of the internal carotid artery. These vessels supply the periorbital area superiorly by the supratrochlear and the supraorbital arteries. The rich anastomoses between these systems ensure survival of random flaps in this area and provide communication between the internal and external carotids at this level. It is the venous drainage in this system that makes orbital infections uniquely dangerous, with the risk of intracranial spread.

Forehead

The forehead vascular supply comes from the temporal, supraorbital, and trochlear arteries, and drainage is from the accompanying veins of those named vessels. The rich interconnecting blood supply of the region permits survival of the whole unit on a single pedicle. Flaps elevated from part of the forehead for nasal reconstruction were first described in the 18th century, accompanying a rediscovery of the "Indian method" dating back to the 6th century. This flap was contrasted to the pedicle flap of Taglicozze described in the 15th century. In a study of cadaveric circulation, Shumrick describes the dominance of the supratrochlear vessels as they exit the orbit 1.7 to 2.2 cm from the midline, passing medial to the eyebrow and piercing the frontalis muscle. They ascend the forehead in a subcutaneous plane 1.5 to 2.0 cm from the midline as symmetrical paramedian vessels. A median forehead flap can be designed safely based on just one vessel.

Nose

The external nasal blood supply arises from the branches of the facial artery: dorsal nasal, columellar, and terminal branch (i.e., the angular artery). Branches of these vessels run superficial to the musculoaponeurotic layer of the nose. Disruption of these vessels may effect open rhinoplasty and supratip debulking and lead to an increase in postoperative nasal edema following these procedures. The intranasal mucosa, bony and cartilagenous septa, and turbinates are supplied by branches of the sphenopalatine artery, which is a terminal branch of the internal maxillary artery. Variations on this vascular supply are not uncommon.

Mandible

The primary blood supply comes from the maxillary artery, via the alveolar branches that arise immediately after the middle meningeal branches. The alveolar artery supplies the lingual and mental branches as well. Fractures of the mandible, which can disrupt the alveolar artery, rarely result in avascular necrosis because the anastomoses between the right and left sides and associated systems provide adequate retrograde blood flow. Postradiation radionecrosis can be a significant problem in this region, especially in the presence of dental caries. Free flap reconstruction of the mandible, or even of the midface region, is generally dependent on the facial artery or other branches of the external carotid located outside the immediate field of surgery, as well as radiation.

Maxilla

The blood supply to the pharynx, tonsillar pillars, and palate and nose forms a plexus that ensures viability of a variety of elective surgical approaches. Vessels of the sphenopalatine complex link directly to the alar branches of the facial artery, as do the branches of the anterior ethmoidal artery. The supply to the oral surface of the palate is derived primarily from the greater palatine artery via the greater palatine foramen at the posterior shelf of the hard palate. The greater palatine artery is a branch of the sphenopalatine artery. The greater palatine artery communicates anteriorly via the incisor foramen with the posterior septal branch of the sphenopalatine artery. This permits the ready survival of multiple types of cleft palate repair, from the bipedical flaps described by von Langenbach in the 19th century, to the double opposing Z-plasties of Furlow in recent years. The multiple branches of the sphenopalatine artery supplying the turbinates and septa of the nose may be sources of epistaxis. Failure of direct local control of a nose bleed, either by cautery or packing, may require endoscopic ligation of the maxillary artery through the maxillary sinus.

Scalp

The scalp is directly supplied by the supraorbital and supratrochlear branches of the internal carotid artery, as well as by the superficial temporal and occipital branches of the external carotid artery. Different flaps for forehead, nasal, and external ear reconstruction have been designed based on these axial vessels. The intercommunications of the system allow extension of the flaps beyond the midline, or the replantation of the whole scalp based on one or preferably two vessels.

CONGENITAL MALFORMATIONS AND CONDITIONS OF THE VASCULAR SYSTEM OF THE HEAD AND NECK

The vascular malformations of the head and neck are hemangiomas, port wine stains, and mixed lymphatic/vascular malformations. The array of classifications and the variability of treatment options and outcomes reported result from confusion regarding these lesions and their correct identification. Mulliken and Glowacki divided the vascular anomalies into two major categories: hemangiomas, which they call tumors, and vascular malformations. This classification is based on the differences and similarities of conditions derived by comparing histology, biochemistry, and clinical differences, and is supported by the extensive work done by Mulliken and his colleagues. The most complete analysis of these differences is found in the published work of Mulliken and Young.

Mulliken identifies hemangiomas (the most common tumor of infancy, occurring in ~12% of infants) as proliferative endothelial cell tumors. These are characterized by their sudden appearance within the first 3 months after birth, rapid growth over several weeks and months, followed by involution and regression beginning at 9 to 10 months, and often continuing for 10 years. Generally, it is safe to predict that the appearance will be satisfactory by age 5 or so, fortunately, corresponding with the social needs of the child.

Vascular malformations are errors of vascular morphogenesis and are classified by the channel abnormalities present. They may be divided into three categories: (1) high flow, such as arteriovenous fistulas (AVFs), which have direct connections between arteries and veins, and arteriovenous malformations (AVMs), with plexiform collections of vessels between feeding arteries and draining veins; (2) low flow, such as capillary malformations (CMs), venous malformation (VMs), and lymphatic malformation (LMs), either macrocystic or microcystic; and (3) combined vascular malformations. These lesions are present at birth, although they may become evident or symptomatic later, grow proportionately to the child, and be stimulated in response to several factors, including hormonal changes of puberty or pregnancy, trauma, including surgery or radiological interventions such as embolizaton, and infection.

High-flow lesions are noted as soft tissue masses involving large areas, crossing soft tissue planes. They may be asymptomatic, with a cutaneous blush, and may or may not have associated soft tissue hypertrophy. Clinically, they may be detected by the presence of local hyperemia, pulsation, an audible bruit, and palpable pulsations. Large lesions in infancy may be associated with high-output cardiac failure. Massive extremity involvement may require amputation. Such lesions of the head and neck (called dental AVMs) are life threatening, with spontaneous hemorrhage associated with tooth eruption, or dental infection or extraction. Ischemic tissue changes can cause pain, ulceration, or hemorrhage.

Surgical intervention of high-flow lesions when symptoms are severe should be preceded by careful evaluation of the region, identifying feeding and draining vessels. Embolization of the lesion may be helpful, but in the head and neck, the rich intercommunications between the extracranial and intracranial circulation make the procedure risky and should be attempted only by experienced interventional radiologists. In their article on interventional radiology, Kagetsu et al (1991) noted the risks to cranial nerve function posed by embolization procedures done to control congenital vascular lesions, epistaxis, or tumor blood supply. They suggested that preembolization injection of xylocaine will determine the extent of the risk and avoid permanent palsy by indicating the need to redirect injection flow. This technique supports the clinical importance of the angiosome concept of Taylor et al (1994), that blood flow is linked to nerve patterns.

Surgical resection of low-flow lesions should be done with careful consideration of the anatomy involved and primarily for cosmetic reasons. Complete resection is not necessarily possible when the lesion is extensive and involving essential structures. Incomplete resection of high- or low-flow lesions may result in recurrence. Surgical management of hemangiomas is generally deferred until well beyond infancy because many of these lesions involute nearly completely. Some authors have suggested recently that cosmetically significant localized lesions can be safely excised once growth has stabilized.

Syndromes associated with vascular malformations of the head and neck include those listed following here.

Sturge-Weber Sequence

This presents with a port wine stain involving the trigeminal nerve distribution, may involve the eye, and includes meningeal and central nervous system (CNS) hemangiomata. Seizures begin by 2 to 7 months. Port wine stains often occur with a facial distribution, but, in the absence of CNS involvement, are not Sturge-Weber.

Von Hippel-Lindau Syndrome

Multiple hemangiomata, beginning principally in the retina and cerebellum, may be associated with other cutaneous (especially the face) and internal organ lesions, multiple cysts of the visceral organs, and organ malignancy (especially renal cell carcinoma) that appears in midadult life.

Peutz-Jeghers Syndrome

Also known as mucocutaneous pigmentation/intestinal polyposis, Peutz-Jeghers syndrome is characterized by pigmented blue or brownish spots appearing in the perioral area, buccal mucosa, and elsewhere during infancy to early childhood. Also characteristic are vascular hamartomas of the stomach and small bowel and occasionally of the nasal oral pharyngeal areas. The syndrome is associated with colonic polyposis, malignancy of the colon, and cancers of the breast, testicle, pancreas, uterus, and ovaries.

Hereditary Hemorrhagic Telangiectasia

Also known as Osler hemorrhagic telangiectasia syndrome, this syndrome is characterized by telangiectases of the tongue, lips, face, conjunctiva, nasal mucous membrane, and fingertips and nail beds. The lesions may be nodular, spider, or pinpoint in nature. Occasional AVM of the lungs and viscera may appear. The most common first indication of this condition is epistaxis.

THE LYMPHATIC SYSTEM

The lymphatic drainage of the head and neck has been defined by proximity of the nodes, in a dependent position to the affected area, and by the anatomical triangles of the neck. The mechanical role of the lymphatic system, as a drainage filter, with adenopathy an indicator of infection or malignancy, is well established. The immune role continues to evolve, with increasing understanding of the local and systemic role these glands offer.

The advent of antibiotics has reduced the frequency of the significant massive lymphadenopthy associated with chronic otitis media, mastoiditis, and tuberculosis (scrofula) seen in the first half of the 20th century. The use of "skinny needle" biopsy techniques has reduced the need for diagnostic open biopsies of lymphadenopathy. The development of sentinel node biopsy techniques has permitted a more specific identification of nodal drainage patterns and has begun to challenge old concepts of cancer surgery. The parotid gland, with its numerous intraglandular lymph nodes, no longer needs to be harvested in continuity with facial malignancies, unless lymph scintigraphy and the lymphazine dye tests indicate the presence of affected nodes. Similarly, complete node dissection may not be necessary as often as once supposed. The application of sentinel node techniques in melanoma is now a well-established practice, and the extension of the method to the treatment of other squamous mucosal and cutaneous tumors is being investigated. The ability to identify the specific first site of lymphatic metastasis will eliminate unnecessary surgery and continues the trend away from "heroic" procedures.

The anatomy of the lymphatic chains follows the venous drainage patterns of the head and neck, with significant concentration along the course of the deep jugular system. Scalp drainage occurs into the posterior occipital region and the posterior triangle. The anterior scalp, forehead, and lateral facial areas drain into the periparotid nodes. Alveolar inflammations and oral infections affect the submandibular and submental nodes. The pharynx is drained by the deep jugular nodes, and more distally, by the mediastinum. The thyroid drains into the jugular chain and mediastinum.

Suggested Readings

Kagetsu NJ, Gerenstein A, Choi IS. Interventional radiology of the extracranial head and neck. Cardiovasc Intervent Radiol 1991;14:325–333

Mullikan JF, Young AE. Vascular Birthmarks. Philadelphia: WB Saunders; 1988

Takahashi K, Mulliken J, Kozakewich H. Cellular markers that distinguish the phases of hemangiomas during infancy and childhood. J Clin Invest 1994;93:2357–2364

Taylor GI, Gianoutsos MP, Morris SF. The neurovascular territories of the skin and muscles: anatomic study and clinical implications. Plast Recon Surg 1994;94:1–36

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. The principle branches of the extracranial internal carotid artery are the
   - A. Superior thyroidal artery
   - B. Lingual artery
   - C. External carotid artery
   - D. None of the above
2. Venous drainage of the head and neck poses the greatest risk of intracranial abcess formation via the
   - A. Internal jugular vein
   - B. Branches between the sphenopalatine system and the meninges
   - C. Communications between the cavernous sinus and the ophthalmic veins
   - D. A and B primarily
3. The terminal branch of the internal maxillary artery is the
   - A. Angular artery
   - B. Sphenopalatine artery
   - C. Infraorbital artery
   - D. Superior pharyngeal artery

Chapter 59
THE BIOLOGY OF FLAPS

NEAL FUTRAN

Anatomical defects created by trauma or tumor excision have necessitated the evolution of reconstructive flap surgery. Flaps are the building blocks that enable both form and function to be restored with safety and reliability. Essential to the survivability and ultimate clinical success of flaps is their ability to sustain sufficient blood flow. It is now generally agreed that the anatomical vascular basis of the flap provides the most accurate approach for classification. Method of movement and flap composition, however, are also used to describe flaps. The objectives of this chapter are to describe the vascular anatomy of flaps, describe the concept of anatomical territories supplied by specific source vessels (the angiosome concept), and discuss flap dynamics. In addition, methods to enhance flap survival will be delineated.

VASCULAR ARCHITECTURE

The vascular architecture of the skin and the deep tissues is an unbroken network of interconnecting vessel arcades. The skin is supplied by three principal vascular sources: the direct cutaneous system of vessels, musculocutaneous perforators, and the fasciocutaneous system. Direct cutaneous arteries arising from segmental and axial vessels constitute the primary skin supply. These lie above the muscular fascia and are parallel to the skin. Whether they follow the intermuscular septae or pierce muscles en route, their main destination is the skin. Indirect vessels constitute the secondary cutaneous supply. They emerge from the deep fascia as terminal branches of arteries whose main purpose is to supply the muscles in other deep tissues. They reinforce and interconnect with the primary supply to the skin. This muscular vasculature produces perforating musculocutaneous branches that enter the subcutaneous tissue, and a rich anastomotic network exists with the direct cutaneous system of vessels. Both of these perforators then arborize in a subfascial and prefascial plane, a subdermal and dermal plane, and a subepidermal plane. Arterioles lie in the subcutaneous tissues and give off branches, which form a network in the deep part of the dermis and on its undersurface. Branches supply the subcutaneous stratum with its fat cells, sweat glands, and the deeper portions of the hair follicles. From the other side of this network vessels enter the dermis from the denser subcapillary network at the junction of the capillary and reticular layers. These give off small branches, which form capillary loops and plexi in the dermal papilla. It is through this pattern of organization that the skin receives a diffuse and overlapping blood supply.

The vessels follow the connective tissue framework of the body. At first they are protected by the bone and cartilaginous framework of the body and then follow the intermuscular septae. They are then distributed to the various tissues along their connective tissue frameworks. The vessels radiate from fixed to mobile areas, and there is a parallel relationship between tissue mobility and the size and density of the supplying vessels. In areas where tissues are mobile, the vessels are large, spaced well apart, and course for long distances parallel to the plane of mobility. The converse applies where tissues are fixed over an area. In a fixed region of a muscle, the vessels tend to be smaller and are sited closer together. Vessels tend to travel with nerves, and vessel size and orientation are the product of tissue differentiation and growth in the area.

Angiosomes or vascular territories are described anatomically as a block of tissue composed of skin and underlying deep structures that are supplied and drained by a named source. When a flap is elevated, angiosomes are captured in sequence on the artery at the base of the flap, each linked by smaller choke vessels. Based on clinical experience, it has been shown that in most cases, when a flap is based on the vessels of one angiosome, the corresponding tissue or tissues of the adjacent angiosome can be included with safety. Distal necrosis will occur in a flap when an attempt is made to include the next, or subsequent, angiosome. This is thought to be due to the pressure gradient that must occur as blood flows across choke vessels that join adjacent arterial territories. The more choke systems there are in a flap, the less pressure is available for flow to the distal end. When multiple angiosomes are traversed, the vascular contributions from each source artery are linked within the muscle, usually by choke vessels. The knowledge of the number and size of these territories within each muscle is essential to flap design. It is one of the factors that explain the reliability of some musculocutaneous flaps and the uncertainty of others.

When a flap based on a specific artery is raised in two or more stages separated by ligation of the adjacent perforator, the delay phenomenon occurs. The choke vessels dilate, and necrosis will occur further along the flap at the next subsequent choke vessel interface. Preexisting vessels between adjacent perforators open up, especially along the axis of the flap. When necrosis of the flap occurs, it is sited usually in the zone of choke vessels that connect adjacent arterial territories. A surgical delay increases the survival length of a flap by allowing the choke vessels to dilate between the adjacent perforators. Surgical delay of one or more pedicles supplying a muscle results in similar changes in the vascular network of that tissue; namely, enlargement of the choke vessels that link vascular territories. This has particular significance in that it may improve the circulation to the tip of a certain muscle in musculocutaneous flaps where the supply is known to be tenuous.

FLAPS USED TO RECONSTRUCT DEFECTS

Flaps used to reconstruct defects are primarily divided into random pattern, axial pattern, musculocutaneous, and fasciocutaneous flaps.

Random Pattern Flaps

Random cutaneous flaps are commonly employed for small defects of the head and neck. The nutrient supply is received from perforating musculocutaneous vessels at the flap base arising from segmental vessels that underlie muscle and subcutaneous tissue. These vessels supply the distal portion of the flap to interconnecting subdermal-subcutaneous and papillary dermis plexi. The maximum survival length of these flaps is directly related to their vessel's perfusion pressures. Although the survival of random flaps is unpredictable clinically, the length to width ratio of 3:1 to 4:1 often provides a viable flap on the face and scalp. If these flaps are delayed, their survival length can be increased by 50 to 100%. Although limited in reach, these flaps provide viable skin and subcutaneous tissue of similar color and texture to close adjacent wounds.

Axial Pattern Flaps

Axial pattern flaps are based on vessels of the direct cutaneous system that consist of certain specific arteries usually accompanied by veins that run in the subcutaneous fat parallel to the skin surface. These are confined to certain specific sites of the body. The anatomical territories of these axial direct cutaneous arteries generally coincide with the limits and ramifications depicted for those vessels in classical anatomical texts. As noted previously, the adjacent angiosome usually can be captured safely. Therefore, axial flaps can be of greater length to width ratios than random pattern flaps. Utilizing a delay procedure may also increase the length. The availability of these flaps has had a tremendous impact on reconstruction, particularly in the head and neck and upper extremity.

Musculocutaneous Flaps

The musculocutaneous flap is a compound flap in which muscle, fascia, subcutaneous fat, and skin are combined as one unit of tissue based on one or more vascular pedicles. Blood supply to these flaps is distributed through three types of vessels. Segmental vessels are large main vessels that course to the muscle and provide branches to the muscle and overlying skin. Perforator vessels supply circulation to the muscles and provide communication

TABLE 59–1 CLASSIFICATION OF MUSCLE FLAPS

	Blood Supply
Type I	Single dominant pedicle
Type II	Dominant pedicle and multiple minor pedicles
Type III	Dual dominant pedicles
Type IV	Multiple segmental pedicles
Type V	One dominant pedicle and several segmental pedicles

between the deeper segmental vessels and cutaneous vessels. Cutaneous vessels may arise directly from segmental vessels and pass through the muscle and subcutaneous tissue to provide circulation to the skin. They may also arise from perforating vessels that supply the muscle. A rich, anastomotic network usually exists between the two types of cutaneous vessels. It is the ability of the skin to survive on the underlying muscle that characterizes a musculocutaneous flap and not the merely random combination of muscle and skin. Mathes and Nahai (1982) have classified of myocutaneous flaps into five different patterns based on size and number of vascular pedicles to the muscle (**Table 59–1**). In clinical practice, these flaps have the advantage of good vascularity and bulk and the potential to create an innervated reconstruction.

FASCIOCUTANEOUS FLAPS

Fasciocutaneous flaps are a unique entity in which the vessels supplying the skin consist of perforators that pass up to the surface, along the fascial septae, between adjacent muscle bellies and then fan out a the level of the deep fascia to form a plexus from which branches are given off to supply overlying subcutaneous tissues and dermis. The presence of this plexus explains why a fasciocutaneous flap can be raised with greater safety in a much greater length to width ratio than an equivalent purely cutaneous flap in the same location. The upper and lower extremities principally supply the source of these flaps. A classification scheme proposed by Cormack and Lamberty (1986) is described in **Table 59–2**.

TABLE 59–2 CLASSIFICATION OF FASCIAL FLAPS

	Blood Supply
Type A	Multiple perforators at base oriented along the long axis of the flap
Type B	Single vascular pedicle feeding the fascial plexus
Type C	Segmental vessels along the length of the flap
Type D	Composite flap comprising bone, muscle, fascia, and skin supplied by a single vascular pedicle

FLAP DYNAMICS

The cutaneous blood flow must fulfill the metabolic needs of the skin and maintain thermal homeostasis requirements of the body. Normal total blood flow in skin is ~20 mL per minute per 100 g. Several conditions contribute to control of blood flow through the flap at both systemic and local levels. Myogenic responsiveness of the precapillary vessels tends to keep blood flow constant through the capillary bed. It also may help control the amount of fluid filtered out of the capillaries and is responsible for cutaneous autoregulation. Vasoconstrictor tone also can be mediated by contracting factors produced from the endothelium. These appear to be released in response to various chemical and physical stimuli.

Cutaneous vessel diameters are influenced by neural control. Neuropeptides produced and released by nerve endings have effects on vascular tone and wound healing. Predominant compounds described include calcitonin gene–related peptide, substance P, and vasoactive intestinal polypeptide. Noradrenaline and other constrictive hormones cause contraction of blood vessels, whereas β–adrenergic stimulation produces relaxation of arteries partly by hyperpolarizing them and the expulsion of the calcium from the cells.

Temperature also has an effect on tissue perfusion in that cooling results in vasoconstriction due to both noradrenaline release and direct action of cold on smooth muscle. Experiments indicate that it is helpful to keep flaps warm (i.e., at their normal temperature) to minimize decrease in blood flow. Microcirculatory intravascular thrombosis with sequestration of platelets and fibrinogen also can inhibit tissue or perfusion. Platelet aggregation, in addition to serving as a physical barrier to blood flow, may release vasoconstrictors, including serotonin and thromboxane A_2. Strategies to enhance flap survival have focused on agents that reduce and inhibit platelet aggregation.

Finally, control of transcapillary fluid exchange is central to homeostasis of the microcirculation. The presence of edema must interfere with the diffusion of nutrients to the cells of the flap and may affect the circulation through it. The four driving pressures for movement of fluid between capillary blood and the surrounding tissues are (1) hydrostatic capillary pressure, which tends to push fluid out of the capillary; (2) oncotic pressure generated by the plasma proteins, which attracts fluid from the interstitium into the capillary; (3) hydrostatic pressure in the tissue fluid, which pushes fluids into the capillary; and (4) oncotic pressure developed by plasma proteins and other colloids in the interstitium, which attracts fluid from the capillaries into that tissue. Variations among these pressures determine the direction and magnitude of fluid exchange across the permeable capillary wall. Control of these fluid exchanges is beyond the scope of this chapter but is necessary to achieve optimal flap nutrition and perfusion.

In addition to changes in tissue blood supply when creating a flap, tension on that flap may alter its survivability. Collagen types I and III provide the principal supporting framework of the extracellular matrix in the dermis. Elastic fibers in combination with collagen provide the elasticity and relative ease of deformation of the skin. The ability to deform the skin mechanically is limited, however. Excessive tension can result in flap necrosis, perhaps caused by reduced blood flow, and should be avoided in the clinical setting.

Several extrinsic disease states play a role in flap survival. Nicotine from cigarette smoke has been shown to decrease skin flap blood flow, cause endothelial damage, and inhibit platelet aggregation in animal models. Microcirculatory changes in patients with diabetes mellitus have been implicated in decreased flap survival, as well as nitric oxide system exchanges in patients with hypertension and atherosclerosis. Although advanced age can cause decreased skin perfusion in response to surface temperature, skin flaps in elderly patients have not been shown clinically to have a deleterious effect on survival. Excessive obesity, however, may result in decreased reliability of the cutaneous territory of muscle and fascia. A thick flap has a reduced arc of rotation, and a wide skin island is recommended to ensure incorporation of perforating vessels between muscle or fascia and overlying skin. Hypothyroidism and hyperthyroidism have not been implicated as conditions that decrease flap survival.

ENHANCEMENT OF FLAP SURVIVAL

Skin flap failure is primarily caused by an impairment of the microcirculatory nutrient flow. Many investigations have focused on identifying pharmacological means to improve flap viability. Skin vessels are thought to be 95% integrated by α-adrenergic receptors, and various agents capable of interfering with the formation of noradrenaline or its action on these receptors have been evaluated experimentally. These include phenoxybenzamine, reserpine, thymoxamine, bretyllium, and phentolamine. Regulation of calcium channels across cell membranes can be dramatically altered by ischemic states. Inhibition of calcium entry into cells, therefore, may prolong ischemic tolerance and limit vascular spasm. Calcium channel blockers such as verapamil, diltiazem, nitroglycerine, and nifedipine have been used as vasodilating agents to enhance microcirculatory blood flow. Prostacyclin has been found to improve flap survival, and the effect is likely mediated through its antiplatelet and vasodilating properties. Pentoxifylline, although not a prostaglandin, increases production of prostacyclin and decreases production of thromboxane A_2, resulting in vasodilatation and decreased platelet aggregation. It also lowers blood viscosity, which may be relevant to the treatment of a failing flap. Dipyridamole modifies platelet function by enhancing prostacyclin effect, and ibuprofen inhibits the synthesis of arachidonic acid metabolites and appears to augment flap survival in rats.

In ischemic portions of flaps there is an increased anaerobic metabolism, and subsequently the production of toxic superoxide (free) radicals is elevated. This can occur in the globally ischemic flap undergoing poor perfusion, the distally ischemic flap, and the flap compromised by underlying hematoma. Free radical "scavengers" and production inhibitors have been used to protect the cellular components from destructive effects of these radicals. They may help preserve the microcirculation and survivability of the flap. Allopurinol is a xanthine oxidase inhibitor and may reduce the formation of surperoxide radicals in skin flaps. Superoxide dismutases and mercaptoproprionylglycine also have been reported to scavenge superoxide radicals and help reduce flap necrosis.

SUMMARY

With knowledge of flap dynamics, the vascular supply of flaps, flap composition, and understanding of the concept of angiosomes, many reliable flaps of both small and large surface area can be raised successfully in one stage. Nevertheless, understanding the delay phenomenon provides an opportunity for studying the biochemical and physiological aspects of flap behavior. In addition, technical improvements in flap elevation and vessel preservation have reduced extrinsic causes of flap failure. Pharmacological vasodilators, as well as the reduction of free radicals, hold promise to improved flap survival.

SUGGESTED READINGS

Cormach GC and Lamberry BG. Cadaver studies of correlation between vessel size and anatomical territory of cutaneous supply. Br J Plast Surg 1986 July;39(3):300–336

Daniel RK, Kerrigan CL. Principles and physiology of skin flap surgery. In: McCarthy JG, ed. Plastic Surgery. Philadelphia: WB Saunders; 1990

Gianoutsos MP, Morris SF. The neurovascular territories of the skin and muscles: anatomic study and clinical implications. Plast Reconstr Surg 1994;94(1):1–36

Mathes S, Nahai F. Clinical applications for muscle and musculocutaneous flaps. St. Louis: CV Mosby;1982:44–48

Toriumi, DM, Larrabee WF Jr. Skin grafts and flaps. In: Papel ID, Nachlas, ND, eds. Facial Plastic and Reconstructive Surgery. St. Louis: Mosby Year Book; 1992:31–44

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. Maintaining a length to width ratio of less than 4:1 is essential for which flap to survive?
   A. Random flap
   B. Axial pattern flap
   C. Musculocutaneous flap
   D. Fasciocutaneous flap
2. Which vessels constitute the primary skin supply?
   A. Musculocutaneous
   B. Fasciocutaneous
   C. Direct cutaneous
   D. Axiocutaneous
3. Elevation of a skin flap results in stimulation and activation of which blood vessel component?
   A. Red blood cells
   B. Adventitia
   C. Endothelium
   D. Eosinophils
4. Allopurinol and superoxide dismutase act as
   A. α-adrenergic antagonists
   B. β-adrenergic agonists
   C. Agents to reduce blood viscosity
   D. Free radical scavengers and production inhibitors

Chapter 60

IMPLANTS IN OTOLARYNGOLOGY

G. Richard Holt

Biocompatibility is central to the study of implants used in the face, head, and neck. No material can be considered for use in augmenting, contouring, or replacing native tissue unless it is acceptable to those tissues on a long-term basis, regardless of suitable consistency, strength, weight, appearance, and tractability. Tissue reactions to implanted materials vary, from frank toxicity and rejection to bland extrusion, as seen with most tympanostomy tubes, to bioacceptability with mild or moderate foreign body reaction or fibrous encapsulation, and finally to biointegration, with native cellular tissue growing right up to the material surface or into its pores.

BIOCOMPATIBILITY

Some materials are intrinsically less inert and more likely to induce a severe inflammatory response than others. Factors that may contribute to this difference are bulk consistency (particulate vs nonparticulate), presence of impurities (especially on the surface), intrinsic toxicity causing ongoing cell death, and peptide or saccharide composition, especially with exposed side groups that may contribute to antigenic recognition. In general, the most inert materials are those composed purely of elements nearest to carbon and calcium on the periodic table because these elements compose the greatest (nonwater) part of the human body.

Materials that are only moderately toxic and nonantigenic or are composed of small particles may incite a chronic inflammatory response that may be granulomatous. Implantation of materials that are nontoxic, nonantigenic, and therefore inert, but are not subject to ingrowth or replacement by normal tissue, may be associated with a mild inflammatory response owing to the surgical trauma of implantation. This may subside reasonably rapidly and be succeeded by an influx of fibrocytes, especially if the implant is mobile and continues to injure local tissue slightly through its mobility. A dense, variably thick, fibrous capsule may form around the implant, analogous to scar tissue replacement of injured tissue. Silicone implants characteristically cause the formation of such a dense capsule to some degree. Two generally successful implants in otolaryngology, hydroxyapatite (calcium and phosphorus) and titanium, hold positions on the periodic table close to carbon; a molecular advantage is afforded them in long-term biocompatibility.

A metallic implant or device may be composed of "pure" metal or, more commonly, may be an alloy of several metals. An alloy is developed to improve the bulk qualities of a pure metal base by adding certain other metals whose properties help achieve greater biocompatibility or better mechanical advantage. The metals most commonly used in facial implantation are titanium and stainless steel; other metals such as chromium, aluminum, cobalt, copper, nickel, and tungsten are included in alloys to improve their properties in the tissues. Titanium and its alloys are generally held to be among the most biocompatible metals in current use, and commercially pure titanium appears to have a better biocompatibility than the most commonly used alloy Ti-6Al-4V (89% titanium, 6% aluminum, and 4% vanadium).

Two physical properties of implant materials are important: porosity and particle size. Optimum biocompatibility is facilitated by tissue growth into the pores of porous material. If the pores are smaller than 50 μ, it is unlikely that tissue will grow into the implant. Therefore, the natural material or the manufactured material should have pore sizes greater than 50 μ in diameter to allow for tissue ingrowth. Likewise, the particle size of the base material is related to the inflammatory response. If the base particles are more than 60 μ in diameter, they cannot be phagocytosed by microphages; if smaller, ingestion can lead to macrophage death, with release of lytic intracellular enzymes and reingestion by other macrophages, initiating a chronic inflammatory response. Degradation of implanted materials is another material-specific tissue response to implantation. This degradation may be random hydrolytic cleavage, which is common to most synthetic polymers, or specific enzymatic attack, which is the fate of most biopolymers, such as collagen (chromic catgut sutures). Even collagen that has been stabilized by inducing cross-linkage is degraded enzymatically.

The implant surface–tissue interface is the most important relationship in biocompatibility in vivo: the fate of the implant depends on the events and processes that take place on or adjacent to its surface. Significantly, the collective properties of the surface determine what sort of host reaction will occur. The unique surface qualities of a material are typified by titanium, which forms a layer of titanium oxide ~100 Å thick immediately after machining, when exposed to air. This layer is vital to osseointegration because it acts as a ceramic and promotes excellent tissue biocompatibility. The layer, in fact, continues to thicken slowly over time of tissue implantation.

As a rule, cells do not adhere directly to the surface of synthetic implanted materials. Some substances present in the extracellular matrix (ECM) form a substrate to which cells may attach to spread and proliferate. ECM proteins such as fibronectin, vitronectin, and possibly some proteoglycans provide this substrate. An interaction with a cell membrane receptor furnishes the linkage for a cellular attachment to an adsorbed ECM protein on an implant surface. This cellular attachment to adsorbed proteins constitutes the primary tissue bioacceptability response. The predominant cell that attaches to the protein layer is the fibroblast. The fibroblast lays down immature collagen on the surface and into the interstices (if porous) of the implant, forming a fibrous capsule or ingrowth of collagen fibers to secure the implant in position.

Surgical technique is another aspect of implantation that affects tissue reaction. After the material to be implanted has been selected, it must be fashioned by the surgeon (unless it comes preformed) into an appropriate size and shape. Implants that are too large for the space in which they are placed put undue tension on overlying tissues and may press against nearby structures such as bone, nerves, or vessels, contributing to possible ischemia, necrosis, or chronic severe inflammation, which could lead ultimately to implant failure. Implants should be shaped according the application, and sharp, jagged, irregular edges should be smoothed. Irregular margins not only may lead to suboptimal aesthetics but also may cause ongoing tissue injury, leading to a chronic inflammatory response.

Once properly located, implants should be fixed in position to avoid migration or mobility with sutures, external taping, or a firm, form-fitting tissue "pocket." Mobility may render an implant more susceptible to external trauma or movement by the patient, which could cause fragmentation and subsequent severe inflammation. Adequate soft tissue coverage is a requisite for optimal long-term implantation; thin, tenuous tissue covering an implant will increase the chances of infection or extrusion, regardless of the environment at the implant–tissue interfaces. Recipient tissues should be handled gently, and efforts should be made to avoid hematoma formation at the implant site.

METALLIC IMPLANTS

Metals are crystalline materials with an orderly three-dimensional arrangement of atoms and molecules. Most metals have a basic cubic or hexagonal shape to form a lattice structure. This lattice structure can be changed to some degree by heating, cooling, hardening, or deformation. Natural defects in the lattice structure exist and can be filled by substitution of other elemental atoms to form alloys, thus strengthening the lattice structure and decreasing its risk of fatigue or fracture. However, most metals will undergo "relaxation" with time, which can lead to fatigue and failure with chronic loading; this may be a problem with mandibular implants, which are subject to the forces of chewing.

Whereas most metals undergo some form of oxidation (corrosion) when exposed to tissue fluid, stainless steel is quite resistant to oxidation due to the presence of other metals in its iron-based alloy composition—chromium, nickel, molybdenum, and manganese. Stainless steel, however, can undergo plastic deformation with time. It is used in reconstruction bars, plates, and fracture repair wires.

Titanium is lightweight, very resistant to corrosion, and very biocompatible. Commercially pure titanium is rather soft, and if not anchored in bone, it can be deformed by loading forces. If the stress response of a metal is too rigid, however, it can lead to stress shielding of the bone, an unwanted condition not conducive to new bone formation kinetics (remodeling). Engineering stress represents the material's ability to handle a given load per cross-sectional area; the loading characteristics of an implant need to be matched with that of the normal maxillofacial structures and their function. The pure form is used in anchoring vibration hearing aids to the skull and prosthetic ears, orbits, and noses to the face.

CERAMICS

Like metal, ceramics have a microlattice structure called "oxides." Glasses, however, have an amorphous, less ordered structure, and are characteristically more brittle and subject to fracture. Most ceramic implants are composed of the base element silica (SiO_2), with other crystalline materials embedded in the noncrystalline glass. Ceramics are thermal resistant and can be used where thermal shock may occur. However, susceptibility to crack propagation is high and is due to stress concentration on the microlattice structure. Clinically, ceramics are brittle materials that will not bend but will fracture with stress and strain. Their main value is a relatively good biocompatibility.

Hydroxyapatite is composed of elements found in the ground substance of bone: calcium and phosphorus. It is capable of osseoconduction of bone cells and, to some degree, osseoinduction in the presence of viable osteoblasts. It can be used in the skull as a resorbable osteogenic bone-densifying implant material; in the mandible as a bone-conductive (regenerative) implant; and as a load-bearing, preformed implant (when combined with a metal) to replace the temporomandibular joint. It is available in powder form, which can be mixed into a paste, or as a preformed, porous implant. Ceramic hydroxyapatite involves the heating of apatite crystals to a high temperature (sintering), which fuses them together in a solid form. Nonceramic hydroxyapatite is nonsintered and is produced by direct crystallization of hydroxyapatite at physiological temperatures and pH. The latter form can be "cured" in the patient while conforming to the bony defect, even in the presence of blood or saline.

POLYMERS

The variation in synthesis techniques can result in widely diverse mechanical properties. Properties of polymers are related to their structure and chemical composition, which in turn are due to the length and type of polymerization, and the degree of cross-linking to form low to high molecular weight polymers. The resulting material can be soft and weak or hard and brittle, depending on the polymerization process used.

Silicone injections and implants were used for soft tissue augmentation for many years; however, their propensity to form dense capsules and the concern about autoimmune disorders arising from polyurethane-covered breast implants have greatly reduced their utilization in favor of newer products. Polymethyl methacrylate was a polymer implant that hardened upon curing; it was used as a replacement for the skull bone. Teflon formerly was injected into a paralyzed vocal cord to improve airway protection and vocalization; however, it caused the formation of granulomata on the focal fold that were very difficult to remove. Prolene (polypropylene) and Vicryl (polylactide/polyglycolide) polymer sutures are commonly used in surgical procedures at this time.

Expanded polytetrafluoroethylene (Gore-Tex) biomaterials are commonly used as soft tissue augmentation implants. This material is woven, with welded nodes, fibers, and microfibers connecting the nodes, and an internodal distance of 10 to 50 μ, which allows for soft tissue ingrowth and stabilization of the implant. It is available in sheets (onlay and suspension grafts of the face), sutures (for long-term suspension of the paralyzed face), tubes (augmentation of the lips), and as preformed implants (ear, cheek, chin, nose). These implants are soft, malleable, and easily sutured to the surrounding tissues for stabilization. Because Gore-Tex is pliable, the implant will move with the overlying skin in a more natural manner. Extrusion is unusual, unless trauma or thin overlying tissue is present.

Polyamide mesh (nylon, Supramid) is best used as a distant implant, allowing fibrovascular ingrowth into the implant, with transposition to the facial region after several weeks. The nylons can undergo hydrolytic degradation with time, leading to partial implant absorption. A new form, polyethylene terephthalate (Mersilene), does not absorb and can be used as a strong suspensory material in the face. High-density polyethylene (Medpore) is a relatively compressible material with good flexibility. Its pore size of greater than 100 μ allows ingrowth by soft tissue or bone. A "Velcro-like" surface keeps the implant securely positioned in the tissues. It is currently used as an ear cartilage replacement framework in children with microtia (failure of growth of the ear).

BIOLOGICAL IMPLANTS

Collagen is the prototypical biological implant. Zyplast and Zyderm are injectable bovine collagen forms that are emulsified and injected intradermally to augment soft tissue defects or to fill in grooves and wrinkles in the face. Because it is an allograft, the patient must be skin-tested before injection to identify possible hypersensitivity. Human fat and dermis can be aspirated from a donor site and reinjected into a facial recipient site for augmentation of depressed scars or atrophied soft tissues. The use of an insulin bath for the fat implant may decrease lipolysis and increase glycogen and lipid formation within the cells. Zyderm is concentrated collagen and requires overcorrection by 150 to 200% to reach a final acceptable level of augmentation. Zyplast is cross-linked with glutaraldehyde to stabilize the collagen, which increases the survival time in human tissues and requires less overcorrection.

Alloderm is cadaver dermis that has been harvested and depleted of any cellular remnants to render it nonantigenic. It can be used in sheets to augment or cover defects, including intraoral ones, or can be injected into lips and scars for augmentation. Because the "ghost vessels" are still present, although the endothelial cells have been evacuated, angioneogenesis can occur from the recipient bed to the graft and reestablish blood flow to the implant. Alloderm requires rehydration in two cycles of 5 minutes each in sterile saline. It is stated that human acellular dermis undergoes no volumetric change nor resorption in the body.

SUMMARY

The soft tissue response to an implanted synthetic material is an inflammatory reaction to a foreign body; factors that minimize this inflammation will maximize biocompatibility and longevity in the tissues. The ideal implant is selected from a material that is nontoxic, nonantigenic, and in chemical proximity to calcium or carbon on the periodic table. If it is porous, the pores should be large enough to admit immune and phagocytic cells and, ideally, to allow native tissue ingrowth. The implant should be of appropriate size and shape and should be implanted in the correct location.

The material should be nonparticulate, should resist fragmentation, and should be secured in the selected location after gentle insertion. All these factors help decrease the body's natural response to an implanted foreign body, but inflammation and foreign body reaction are the common threads in all responses to all implanted synthetic materials. Optimum soft tissue or osseous biocompatibility, characterized by the stabilizing ingrowth of mesenchymal tissue into the pores and interstices, is achieved by formulating and fabricating the material properly and by avoiding or containing the inflammatory response.

Commonly used implant materials for the face, head, and neck include biological materials such as collagen, ceramics, polymers, metals, and autologous tissues. The microstructure and surface characteristics of the implant are important in determining the recipient site's response to the implanted material. In bone, fixation must be firm, but it must allow for micromotion to reduce stress shielding of bone remodeling. Soft tissue augmentation or suspensory materials must have some pliability to "move" with facial motions and to be soft enough not to be overly noticeable beneath the skin.

SUGGESTED READINGS

Hom DB. Contouring with facial implants. Adv Otolaryngol Head Neck Surg 1997;11:285–311

Morehead JM, Holt GR. Soft-tissue response to synthetic biomaterials. Otolaryngol Clin North Am 1994;27:195–201

Rubin JP, Yaremchuk MJ. Complications and toxicities of implantable biomaterials used in facial reconstructive and aesthetic surgery: a comprehensive review of the literature. Plast Reconstr Surg 1997;100:1336–1353

Sclafani AP, Thomas JR, Cox AJ, Cooper MH. Clinical and histologic response of subcutaneous expanded polytetrafluoroethylene (Gore-Tex®) and porous high-density polyethylene (Medpor®) implants to acute and early infection. Arch Otolaryngol Head Neck Surg 1997;123:328–336

Vacanti CA, Vacanti JP. Bone and cartilage reconstruction with tissue engineering approaches. Otolaryngol Clin North Am 1994;27:263–276

Weisberger EC, Eppley BL. Resorbable fixation plates in head and neck surgery. Laryngoscope 1997;107:716–719

Self-Test Questions

For each question select the correct answer from the lettered alternatives that follow. To check your answers, see Answers to Self-Tests on page 718.

1. Which of the following is a factor that can contribute to an inflammatory response after implantation of an alloplastic material?
   - A. Less inert
   - B. Bulk consistency
   - C. Surface impurities
   - D. Peptide composition
   - E. All of the above
2. Macrophages cannot permeate materials with a pore size of less than
   - A. 1 μ
   - B. 15 μ
   - C. 25 μ
   - D. 50 μ
   - E. 100 μ
3. Implantation of materials that are nontoxic, nonantigenic, and inert may be associated with a mild inflammatory response owing to
   - A. Surgical trauma of implantation
   - B. Type IV hypersensitivity
   - C. Mobility of the implant
   - D. A and B
   - E. A and C
4. Metals are crystalline materials with a
   - A. Disorderly arrangement of atoms
   - B. Two-dimensional framework
   - C. Cylindrical lattice structure
   - D. Cubic lattice structure
   - E. Resistance to fatigue and failure.
5. Which of the following best describes ceramic implants?
   - A. Most are composed of carbon atoms in array.
   - B. They are never used for load-bearing surfaces.
   - C. They are thermal sensitive.
   - D. Ceramic implants are resistant to crack propagation.
   - E. They can be considered bone conductive.

This page intentionally left blank

ANSWERS TO SELF-TESTS

PART I THE BASIC PRINCIPLES

Chapter 1

1. a, 2. b, 3. c

Chapter 2

1. b This is not a critical step, and healing proceeds normally in its absence.
2. c The first type of cell to migrate to the wound is the neutrophil. Platelets are passive, macrophages respond after 48 hours, and fibroblasts do not predominate until ~5 days have elapsed. Neutrophils are the first responders to the inflammatory kinins produced as a result of an acute injury.
3. d The fibroblast produces collagen as well as the listed ground substance substrates.
4. d Keloids form when there is a propensity for them, regardless of the angle of incision. Relaxed lines of skin tension follow underlying muscle and collagen forces so that incisions will not have widening forces applied. This does not guarantee that hypertrophy will not occur, but it is most likely to provide the best result of healing with satisfactory scar formation.

Chapter 3

1. d, 2. a, 3. c, 4. d

Chapter 4

1. d, 2. c, 3. c, 4. c

Chapter 5

1. a and d
2. b; $VO_2 = Q(CaO_2 - CvO_2)$

$$\begin{aligned} Q &= VO_2/(CaO_2 - CvO_2) \\ &= 400\ \text{mL/minute}/(19\ \text{mL/minute} - 16\ \text{mL/minute}) \\ &= 13.33\ \text{L/minute} \end{aligned}$$

3. a

Chapter 6

1. d, 2. a, 3. b

Chapter 7

1. d, 2. d, 3. e, 4. b, 5. e

Chapter 8

1. a, 2. a, 3. b, 4. d, 5. b, 6. c,
7. d, 8. b, 9. a, 10. b, 11. a, 12. e

Chapter 9

1. c, 2. a, 3. c

Chapter 10A

1. d, 2. d, 3. a, 4. e, 5. a

Chapter 10B

1. e, 2. e, 3. b, 4. d, 5. c

Chapter 11

1. c, 2. d, 3. b, 4. d, 5. d

Chapter 12

1. e, 2. d, 3. d

Chapter 13

1. c, 2. c, 3. b

Chapter 14

1. a, c, f, 2. a, b, 3. b, c, d,
4. a, d, f, 5. a, e, f

Chapter 15

1. c, 2. c, 3. b

Chapter 16

1. b, 2. b, 3. d, 4. d, 5. c

Chapter 17

1. c, 2. e, 3. b, 4. b

Chapter 18

1. c, 2. e, 3. d

Chapter 19

1. c, 2. d, 3. c, 4. c, 5. a, 6. b

PART II THE EAR, HEARING, AND BALANCE

Chapter 20

1. c, 2. e, 3. c, 4. d, 5. a

Chapter 21

1. d, 2. d, 3. c, 4. a and c

Chapter 22

1. d, 2. c, 3. d, 4. a

Chapter 23

1. c, 2. a, 3. d

Chapter 24

1. c, 2. d, 3 c, 4. b, 5. a

Chapter 25

1. c, 2. e, 3. b, 4. c

Chapter 26

1. d, 2. a, 3. a

Chapter 27

1. d The question is asking at what level the ascending auditory pathways become bilateral.
2. a
3. a The unopposed input from the left labyrinth will give the patient the illusion that he or she is rotating from right to left, causing the eyes to deviate (slow phase) to the right; the fast phase of the nystagmus will therefore be to the left. Also, in attempting to correct for the illusory rotation from right to left, the patient will tend to fall to the right.
4. c

Chapter 28

1. b
2. d
3. b BAEPs assess the auditory pathways from the ear to the mesencephalon, but they are not reliable indicators of pathology at more rostral sites.
4. a

Chapter 29

1. b, 2. c, 3. c

Chapter 30

1. b, 2. a, 3. b, 4. c,
5. d, 6. d, 7. b, 8. b

Chapter 31

1. d, 2. c, 3. d, 4. e

Chapter 32

1. e, 2. e, 3. f, 4. d

Chapter 33

1. d, 2. b, 3. a

Chapter 34

1. d, 2. d, 3. c

Chapter 35

1. b, 2. c, 3. d, 4. d, 5. a,
6. d, 7. c, 8. a, 9. d

Chapter 36

1. e, 2. b, 3. d

PART III THE NOSE, OLFACTION, AND THE SINUSES

Chapter 37

1. a, 2. d, 3. a, 4. d, 5. b

Chapter 38

1. d, 2. b, 3. c, 4. b, 5. d, 6. e

Chapter 39

1. d, 2. e, 3. d, 4. e, 5. d

Chapter 40

1. d, 2. a, 3. a, 4. d

PART IV THE LARYNX, VOICE, AND NECK

Chapter 41

1. d, 2. c, 3. b, 4. c

Chapter 42

1. a, 2. b, 3. c, 4. d, 5. e

Chapter 43

1. f, 2. a, 3. b, 4. e

Chapter 44

1. e These are all critical laryngeal functions.
2. a Each cricoarytenoid joint does have an elliptical facet, but it is only 6 mm.
3. a The minimal absolute subglottic air pressure required to set the vocal folds moving varies from 3 cm H_2O in modal range to 6 cm H_2O in head range.
4. c The vocalis portion of the thyroarytenoid is a specialized muscle, which parallels the vocal ligament. Upon contraction it shortens the vocal fold, thereby reducing the tension of the cover and lowering pitch.
5. c In pressed phonation the glottal flow waveforms are smaller than for normal phonation; similarly, within the phonatory cycle the amount of time during which there is an open glottis is less than in normal phonation. Spectrally, unlike breathy voice, where energy is primarily seen in the lower partials, the energy is reduced in the lower partials in pressed phonation.
6. e Doubling of the difference should raise the source acoustic power by ~6 dB.
7. d Sulcus vocalis is an unusual condition in which there is a cul-de-sac or pocket within the epithelial lining, which sticks down to the vocal ligament. There is generally an inflammatory response in the adjacent ligament. The vocal fold is thereby lacking in mucosa and stiff, leading to a very poor, often high-pitched voice with short phonation time and generally with secondary abnormal voicing (hyperfunction of the false vocal folds, etc.).

Chapter 45

1. c, 2. b, 3. e

Chapter 46

1. b, 2. e, 3. d

Chapter 47

1. d, 2. c, 3. b, 4. b

Chapter 48

1. d, 2. d, 3. a, 4. b

Chapter 49

1. d, 2. c, 3. c

Chapter 50

1. **False**. The glossopharyngeal nerve is almost never encountered during neck dissection.
2. **False**. There is great variability in the interrelationships of the branches of the external carotid artery.
3. **True**. The current level-based classification scheme has both clinical and anatomical representations to allow uniformity in classifying nodal disease, both preoperatively and intraoperatively.
4. **True**. Although there is some variability, all lower cranial nerves encountered during routine neck surgery have fairly consistent points in their course that will allow their localization with limited dissection.

Chapter 51

1. d, 2. b, 3. a

PART V: THE ORAL CAVITY, TASTE AND THE GLANDS OF THE NECK

Chapter 52

1. a, b, d
2. a, b, c, d
3. a, b, c, d
4. a, b
5. a, b, d
6. b
7. b, d
8. a, b

Chapter 53

1. d, 2. a

Chapter 54

1. b, 2. b, 3. a, 4. b, 5. e

Chapter 55

1. e, 2. a, 3. d

PART VI: FACIAL PLASTICS AND MISCELLANEOUS

Chapter 56

1. a, 2. b, 3. c, 4. a

Chapter 57

1. a, 2. b, 3. c

Chapter 58

1. c There are no branches of the extracranial portion of the internal carotid artery.
2. c It is through this interconnection that orbital cellulitis can progress to a cavernous vein thrombosis and central nervous system abscess.
3. b

Chapter 59

1. a, 2. c, 3. c, 4. d

Chapter 60

1. e, 2. d, 3. e, 4. d, 5. e

Index

Page numbers followed by *f* or *t* denote figures or tables, respectively.

J

W

X