Harrison's Principles of Internal Medicine 21e

Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson

Show Chapters Hide Chapters

— PART 1: The Profession of Medicine

Updated! Chapter 1: The Practice of Medicine

Updated! Chapter 2: Promoting Good Health

Updated! Chapter 3: Vaccine Opposition and Hesitancy

Updated! Chapter 4: Decision-Making in Clinical Medicine

Updated! Chapter 5: Precision Medicine and Clinical Care

New! Chapter 6: Screening and Prevention of Disease

Updated! Chapter 7: Global Diversity of Health System Financing and Delivery

Updated! Chapter 8: The Safety and Quality of Health Care

New! Chapter 9: Diagnosis: Reducing Errors and Improving Quality

Updated! Chapter 11: Ethical Issues in Clinical Medicine

New! Chapter 12: Palliative and End-of-Life Care

— PART 2: Cardinal Manifestations and Presentation of Diseases

Section 1: Pain

New! Chapter 13: Pain: Pathophysiology and Management

New! Chapter 17: Back and Neck Pain

Section 2: Fever

New! Chapter 18: Fever

New! Chapter 20: Fever of Unknown Origin

Section 3: Nervous System Dysfunction

New! Chapter 22: Dizziness and Vertigo

New! Chapter 23: Fatigue

New! Chapter 24: Neurologic Causes of Weakness and Paralysis

New! Chapter 25: Numbness, Tingling, and Sensory Loss

New! Chapter 26: Gait Disorders, Imbalance, and Falls

New! Chapter 27: Confusion and Delirium

New! Chapter 28: Coma

New! Chapter 29: Dementia

New! Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders

New! Chapter 31: Sleep Disorders

Section 4: Disorders of Eyes, Ears, Nose, and Throat

New! Chapter 32: Disorders of the Eye

New! Chapter 34: Disorders of Hearing

New! Chapter 35: Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat

New! Chapter 36: Oral Manifestations of Disease

Section 5: Alterations in Circulatory and Respiratory Functions

New! Chapter 38: Cough

New! Chapter 39: Hemoptysis

New! Chapter 40: Hypoxia and Cyanosis

New! Chapter 41: Edema

New! Chapter 43: Palpitations

Section 6: Alterations in Gastrointestinal Function

New! Chapter 46: Diarrhea and Constipation

New! Chapter 47: Unintentional Weight Loss

New! Chapter 48: Gastrointestinal Bleeding

New! Chapter 49: Jaundice

New! Chapter 50: Abdominal Swelling and Ascites

Section 7: Alterations in Renal and Urinary Tract Function

New! Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome

New! Chapter 52: Azotemia and Urinary Abnormalities

New! Chapter 54: Hypercalcemia and Hypocalcemia

New! Chapter 55: Acidosis and Alkalosis

Section 8: Alterations in the Skin

New! Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders

New! Chapter 58: Skin Manifestations of Internal Disease

New! Chapter 60: Cutaneous Drug Reactions

New! Chapter 61: Photosensitivity and Other Reactions to Sunlight

Section 9: Hematologic Alterations

New! Chapter 65: Bleeding and Thrombosis

New! Chapter 66: Enlargement of Lymph Nodes and Spleen

— PART 3: Pharmacology

New! Chapter 67: Principles of Clinical Pharmacology

Updated! Chapter 68: Pharmacogenomics

— PART 4: Oncology and Hematology

Section 1: Neoplastic Disorders

New! Chapter 77: Head and Neck Cancer

New! Chapter 85: Renal Cell Carcinoma

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 1: The Practice of Medicine

The Editors

ENDURING VALUES OF THE MEDICAL PROFESSION

> *No greater opportunity, responsibility, or obligation can fall to the lot of a human being than to become a physician. In the care of the suffering, [the physician] needs technical skill, scientific knowledge, and human understanding. Tact, sympathy, and understanding are expected of the physician, for the patient is no mere collection of symptoms, signs, disordered functions, damaged organs, and disturbed emotions. [The patient] is human, fearful, and hopeful, seeking relief, help, and reassurance.*
>
> —*Harrison's Principles of Internal Medicine,* 1950

The practice of medicine has changed in significant ways since the first edition of this book was published in 1950. The advent of molecular genetics, sophisticated new imaging techniques, robotics, and advances in bioinformatics and information technology have contributed to an explosion of scientific information that has changed fundamentally the way physicians define, diagnose, treat, and attempt to prevent disease. This growth of scientific knowledge continues to evolve at an accelerated pace.

The widespread use of electronic medical records and the Internet have altered the way physicians and other health care providers access and exchange information as a routine part of medical education and practice **(Fig. 1-1)**. As today's physicians strive to integrate an ever-expanding body of scientific knowledge into everyday practice, it is critically important to remember two key principles: first, the ultimate goal of medicine is to prevent disease and, when it occurs, to diagnose it early and provide effective treatment; and second, despite 70 years of scientific advances since the first edition of this text, a trusting relationship between physician and patient still lies at the heart of effective patient care.

FIGURE 1-1

The Doctor **by Luke Fildes depicts the caring relationship** between this Victorian physician and a very ill child. Painted in 1891, the painting reflects the death of the painter's young son from typhoid fever and was intended to reflect the compassionate care provided by the physician even when his tools were not able to influence the course of disease. *(Source: History and Art Collection/Alamy Stock Photo.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

biochemistry, cell biology, immunology, and genomics, coupled with newly developed imaging techniques, provide a window into the most remote



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

THE SCIENCE AND ART OF MEDICINE

Deductive reasoning and applied technology form the foundation for the approach and solution to many clinical problems. Extraordinary advances in biochemistry, cell biology, immunology, and genomics, coupled with newly developed imaging techniques, provide a window into the most remote recesses of the body and allow access to the innermost parts of the cell. Revelations about the nature of genes and single cells have opened a portal for formulating a new molecular basis for the physiology of systems. Researchers are deciphering the complex mechanisms by which genes are regulated, and increasingly, physicians are learning how subtle changes in many different genes, acting in an integrative contextual way, can affect the function of cells and organisms. Clinicians have developed a new appreciation of the role of stem cells in normal tissue function, in the development of cancer and other disorders, and in the treatment of certain diseases. Entirely new areas of research, including studies of the human microbiome, epigenetics, and noncoding RNAs as regulatory features of the genome, have become important for understanding both health and disease. Information technology enables the interrogation of medical records from millions of individuals, yielding new insights into the etiology, characteristics, prognosis, and stratification of many diseases. With the increasing availability of very large data sets ("big data") from omic analyses and the electronic medical record, there is now a growing need for machine learning and artificial intelligence for unbiased analyses that enhance clinical predictive accuracy. The knowledge gleaned from the ***science of medicine*** continues to enhance the understanding by physicians of complex pathologic processes and to provide new approaches to disease prevention, diagnosis, and treatment. With continued refinement of unique omic signatures coupled with nuanced clinical pathophenotypes, the profession moves ever closer to practical precision medicine. Yet, skill in the most sophisticated applications of laboratory technology and in the use of the latest therapeutic modality alone does not make a good physician. Extraordinary advances in vaccine platform technology and the use of cryo-electron microscopy for the structure-based design of vaccine immunogens have transformed the field of vaccinology, resulting in the unprecedented speed and success with which COVID-19 vaccines were developed.

When a patient poses challenging clinical problems, an effective physician must be able to identify the crucial elements in a complex history and physical examination; order the appropriate laboratory, imaging, and diagnostic tests; and extract the key results from densely populated computer screens to determine whether to treat or to "watch." As the number of tests increases, so does the likelihood that some incidental finding, completely unrelated to the clinical problem at hand, will be uncovered. Deciding whether a clinical clue is worth pursuing or should be dismissed as a "red herring" and weighing whether a proposed test, preventive measure, or treatment entails a greater risk than the disease itself are essential judgments that a skilled clinician must make many times each day. This combination of medical knowledge, intuition, experience, and judgment defines the ***art of medicine***, which is as necessary to the practice of medicine and the precision medicine of the future as is a sound scientific base, and as important for contemporary medical practice as it has been in earlier eras.

CLINICAL SKILLS

History-Taking

The recorded history of an illness should include all the facts of medical significance in the life of the patient. Recent events should be given the most

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

underlying disease that might not be mentioned in the patient's narrative. In general, patients who feel comfortable with the physician will offer more

The recorded history of an illness should include all the facts of medical significance in the life of the patient. Recent events should be given the most attention. Patients should, at some early point, have the opportunity to tell their own story of the illness without frequent interruption and, when appropriate, should receive expressions of interest, encouragement, and empathy from the physician. Any event related by a patient, however trivial or seemingly irrelevant, may provide the key to solving the medical problem. A methodical review of systems is important to elicit features of an underlying disease that might not be mentioned in the patient's narrative. In general, patients who feel comfortable with the physician will offer more complete information; thus, putting the patient at ease contributes substantially to obtaining an adequate history.

An informative history is more than eliciting an orderly listing of symptoms. By listening to patients and noting the ways in which they describe their symptoms, physicians can gain valuable insight. Inflections of voice, facial expression, gestures, and attitude (i.e., "body language") may offer important clues to patients' perception of and reaction to their symptoms. Because patients vary considerably in their medical sophistication and ability to recall facts, the reported medical history should be corroborated whenever possible. The social history also can provide important insights into the types of diseases that should be considered and can identify practical considerations for subsequent management. The family history not only identifies rare genetic disorders or common exposures, but often reveals risk factors for common disorders, such as coronary heart disease, hypertension, autoimmunity, and asthma. A thorough family history may require input from multiple relatives to ensure completeness and accuracy. An experienced clinician can usually formulate a relevant differential diagnosis from the history alone, using the physical examination and diagnostic tests to narrow the list or reveal unexpected findings that lead to more focused inquiry.

The very act of eliciting the history provides the physician with an opportunity to establish or enhance a unique bond that can form the basis for a good patient–physician relationship. This process helps the physician develop an appreciation of the patient's view of the illness, the patient's expectations of the physician and the health care system, and the financial and social implications of the illness for the patient. Although current health care settings may impose time constraints on patient visits, it is important not to rush the encounter. A hurried approach may lead patients to believe that what they are relating is not of importance to the physician, and, as a result, they may withhold relevant information. The confidentiality of the patient–physician relationship cannot be overemphasized.

Physical Examination

The purpose of the physical examination is to identify physical signs of disease. The significance of these objective indications of disease is enhanced when they confirm a functional or structural change already suggested by the patient's history. At times, however, physical signs may be the only evidence of disease and may not have been suggested by the history.

The physical examination should be methodical and thorough, with consideration given to the patient's comfort and modesty. Although attention is often directed by the history to the diseased organ or part of the body, the examination of a new patient must extend from head to toe in an objective search for abnormalities. The results of the examination, like the details of the history, should be recorded at the time they are elicited—not hours later, when they are subject to the distortions of memory. Physical examination skills should be learned under direct observation of experienced clinicians. Even highly experienced clinicians can benefit from ongoing coaching and feedback. Simulation laboratories and standardized patients play an increasingly important role in the development of clinical skills. Although the skills of physical diagnosis are acquired with experience, it is not merely technique that determines success in identifying signs of disease. The detection of a few scattered petechiae, a faint diastolic murmur, or a small mass in the abdomen is not a question of keener eyes and ears or more sensitive fingers, but of a mind alert to those findings. Because physical findings can change with time, the physical examination should be repeated as frequently as the clinical situation warrants.

Given the many highly sensitive diagnostic tests now available (particularly imaging techniques), it may be tempting to place less emphasis on the physical examination. Some are critical of physical diagnosis based on perceived low levels of specificity and sensitivity. Indeed, many patients are seen by consultants only after a series of diagnostic tests have been performed and the results are known. This fact should not deter the physician from performing a thorough physical examination since important clinical findings may have escaped detection by diagnostic tests. Especially important, a thorough and thoughtful physical examination may render a laboratory finding unimportant (i.e., certain echocardiographic regurgitant lesions). The act of a hands-on examination of the patient also offers an opportunity for communication and may have reassuring effects that foster the patient–physician relationship.

Diagnostic Studies

Physicians rely increasingly on a wide array of laboratory and imaging tests to make diagnoses and ultimately to solve clinical problems; however, such information does not relieve the physician from the responsibility of carefully observing and examining the patient. It is also essential to appreciate the

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Single laboratory tests are rarely ordered. Instead, physicians generally request "batteries" of multiple tests, which often prove useful and can be

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Physicians rely increasingly on a wide array of laboratory and imaging tests to make diagnoses and ultimately to solve clinical problems; however, such information does not relieve the physician from the responsibility of carefully observing and examining the patient. It is also essential to appreciate the limitations of diagnostic tests. By virtue of their apparent precision, these tests often gain an aura of certainty regardless of the fallibility of the tests themselves, the instruments used in the tests, and the individuals performing or interpreting the tests. Physicians must weigh the expense involved in laboratory procedures against the value of the information these procedures are likely to provide.

Single laboratory tests are rarely ordered. Instead, physicians generally request "batteries" of multiple tests, which often prove useful and can be performed with a single specimen at relatively low cost. For example, abnormalities of hepatic function may provide the clue to nonspecific symptoms such as generalized weakness and increased fatigability, suggesting a diagnosis of chronic liver disease. Sometimes a single abnormality, such as an elevated serum calcium level, points to a particular disease, such as hyperparathyroidism.

The thoughtful use of screening tests (e.g., measurement of low-density lipoprotein cholesterol) may allow early intervention to prevent disease (Chap. 6). Screening tests are most informative when they are directed toward common diseases and when their results indicate whether other potentially useful—but often costly—tests or interventions are needed. On the one hand, biochemical measurements, together with simple laboratory determinations such as routine serum chemistries, blood counts, and urinalysis, often provide a major clue to the presence of a pathologic process. On the other hand, the physician must learn to evaluate occasional screening-test abnormalities that do not necessarily connote significant disease. An in-depth workup after the report of an isolated laboratory abnormality in a person who is otherwise well is often wasteful and unproductive. Because so many tests are performed routinely for screening purposes, it is not unusual for one or two values to be slightly abnormal. Nevertheless, even if there is no reason to suspect an underlying illness, tests yielding abnormal results ordinarily are repeated to rule out laboratory error. If an abnormality is confirmed, it is important to consider its potential significance in the context of the patient's condition and other test results.

There is almost continual development of technically improved imaging studies with greater sensitivity and specificity. These tests provide remarkably detailed anatomic information that can be pivotal in informing medical decision-making. MRI, CT, ultrasonography, a variety of isotopic scans, and positron emission tomography (PET) have supplanted older, more invasive approaches and opened new diagnostic vistas. In light of their capabilities and the rapidity with which they can lead to a diagnosis, it is tempting to order a battery of imaging studies. All physicians have had experiences in which imaging studies revealed findings that led to an unexpected diagnosis. Nonetheless, patients must endure each of these tests, and the added cost of unnecessary testing is substantial. Furthermore, investigation of an unexpected abnormal finding may lead to an iatrogenic complication or to the diagnosis of an irrelevant or incidental problem. A skilled physician must learn to use these powerful diagnostic tools judiciously, always considering whether the results will alter management and benefit the patient.

MANAGEMENT OF PATIENT CARE

Team-Based Care

Medical practice has long involved teams, particularly physicians working with nurses and, more recently, with physician assistants and nurse practitioners. Advances in medicine have increased our ability to manage very complex clinical situations (e.g., intensive care units [ICUs], bone marrow transplantation) and have shifted the burden of disease toward chronic illnesses. Because an individual patient may have multiple chronic diseases, he or she may be cared for by several specialists as well as a primary care physician. In the inpatient setting, care may involve multiple consultants along with the primary admitting physician. Communication through the medical record is necessary but not sufficient, particularly when patients have complex medical problems or when difficult decisions need to be made about the optimal management plan. Physicians should optimally meet face-to-face or by phone to ensure clear communication and thoughtful planning. It is important to note that patients often receive or perceive different messages from various care providers; thus, attempts should be made to provide consistency among these messages to the patient. Management plans and treatment options should be outlined succinctly and clearly for the patient.

Another dimension of team-based care involves allied health professions. It is not unusual for a hospitalized patient to encounter physical therapists, pharmacists, respiratory therapists, radiology technicians, social workers, dieticians, and transport personnel (among others) in addition to physicians and nurses. Each of these individuals contributes to clinical care as well as to the patient's experience with the health care system. In the outpatient setting, disease screening and chronic disease management are often carried out by nurses, physician assistants, or other allied health professionals.

The growth of team-based care has important implications for medical culture, student and resident training, and the organization of health care systems. Despite diversity in training, skills, and responsibilities among health care professionals, common values need to be espoused and reinforced. Many medical schools have incorporated interprofessional teamwork into their curricula. Effective communication is inevitably the most

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

systems. Despite diversity in training, skills, and responsibilities among health care professionals, common values need to be espoused and reinforced. Many medical schools have incorporated interprofessional teamwork into their curricula. Effective communication is inevitably the most challenging aspect of implementing team-based care. While communication can be aided by electronic devices, including medical records, apps, or text messages, it is vitally important to balance efficiency with taking the necessary time to speak directly with colleagues.

The Dichotomy of Inpatient and Outpatient Internal Medicine

The hospital environment has undergone sweeping changes over the past few decades. Emergency departments and critical care units have evolved to manage critically ill patients, allowing them to survive formerly fatal conditions. In parallel, there is increasing pressure to reduce the length of stay in the hospital and to manage complex disorders in the outpatient setting. This transition has been driven not only by efforts to reduce costs but also by the availability of new outpatient technologies, such as imaging and percutaneous infusion catheters for long-term antibiotics or nutrition, minimally invasive surgical procedures, and evidence that outcomes often are improved by reducing inpatient hospitalization.

In addition to traditional medical beds, hospitals now encompass multiple distinct levels of care, such as the emergency department, procedure rooms, overnight observation units, critical care units, and palliative care units. A consequence of this differentiation has been the emergence of new specialties (e.g., emergency medicine and end-of-life care) and the provision of in-hospital care by hospitalists and intensivists. Most *hospitalists* are board-certified internists who bear primary responsibility for the care of hospitalized patients and whose work is limited entirely to the hospital setting. The shortened length of hospital stay means that most patients receive only acute care while hospitalized; the increased complexities of inpatient medicine make the presence of an internist with specific training, skills, and experience in the hospital environment extremely beneficial. *Intensivists* are board-certified physicians who are further certified in critical care medicine and who direct and provide care for very ill patients in critical care units. Clearly, an important challenge in internal medicine today is to ensure the continuity of communication and information flow between a patient's primary care physician and those who are in charge of the patient's hospital care. Maintaining these channels of communication is frequently complicated by patient "handoffs"—i.e., transitions from the outpatient to the inpatient environment, from the critical care unit to a general medicine floor, from a medical to a surgical service and vice versa, from the hospital environment to the recently developed "home hospital" setting (for select patients with adequate home support), and from the hospital or home hospital to the outpatient environment.

The involvement of many care providers in conjunction with these transitions can threaten the traditional one-to-one relationship between patient and primary care physician. Of course, patients can benefit greatly from effective collaboration among a number of health care professionals; however, *it is the duty of the patient's principal or primary physician to provide cohesive guidance through an illness.* To meet this challenge, primary care physicians must be familiar with the techniques, skills, and objectives of specialist physicians and allied health professionals who care for their patients in the hospital. In addition, primary care physicians must ensure that their patients benefit from scientific advances and the expertise of specialists, both in and out of the hospital. Primary care physicians should explain the role of these specialists to reassure patients that they are in the hands of physicians best trained to manage their current illness. However, the primary care physician should assure patients and their families that decisions are being made in consultation with these specialists. The evolving concept of the "medical home" incorporates team-based primary care with subspecialty care in a cohesive environment that ensures smooth transitions of care.

Mitigating the Stress of Acute Illness

Few people are prepared for a new diagnosis of cancer or anticipate the occurrence of a myocardial infarction, stroke, or major accident. The care of a frightened or distraught patient is confounded by these understandable responses to life-threatening events. The physician and other health providers can reduce the shock of life-changing events by providing information in a clear, calm, consistent, and reassuring manner. Often, information and reassurance need to be repeated. Caregivers should also recognize that, for the typical patient, hospital emergency rooms, operating rooms, ICUs, and general medical floors represent an intimidating environment. Hospitalized patients find themselves surrounded by air jets, buttons, and glaring lights; invaded by tubes and wires; and beset by the numerous members of the health care team—hospitalists, specialists, nurses, nurses' aides, physician assistants, social workers, technologists, physical therapists, medical students, house officers, attending and consulting physicians, and many others. They may be transported to special laboratories and imaging facilities replete with blinking lights, strange sounds, and unfamiliar personnel; they may be left unattended at times; and they may be obligated to share a room with other patients who have their own health problems. It is little wonder that patients may find this environment bewildering and stressful. The additive effects of an acute illness, unfamiliar environment, multiple medications, and sleep deprivation can lead to confusion or delirium, especially in older hospitalized patients. Physicians who appreciate the hospital experience from the patient's perspective and who make an effort to guide the patient through this experience may make a stressful situation more tolerable and enhance the patient's chances for an optimal recovery.

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

decision-making process involves the ordering of additional tests, requests for consultations, decisions about treatment, and predictions concerning prognosis. This process requires an in-depth understanding of the pathophysiology and natural history of disease. Formulating a differential

Medical Decision-Making

Medical decision-making is a fundamental responsibility of the physician and occurs at each stage of the diagnostic and therapeutic process. The decision-making process involves the ordering of additional tests, requests for consultations, decisions about treatment, and predictions concerning prognosis. This process requires an in-depth understanding of the pathophysiology and natural history of disease. Formulating a differential diagnosis requires not only a broad knowledge base but also the ability to assess the relative probabilities of various diseases for a given patient. Application of the scientific method, including hypothesis formulation and data collection, is essential to the process of accepting or rejecting a particular diagnosis. Analysis of the differential diagnosis is an iterative process. As new information or test results are acquired, the group of disease processes being considered can be contracted or expanded appropriately. Whenever possible, decisions should be evidence-based, taking advantage of rigorously designed clinical trials or objective comparisons of different diagnostic tests. *Evidence-based medicine* stands in sharp contrast to anecdotal experience, which is often biased. Unless attuned to the importance of using larger, objective studies for making decisions, even the most experienced physicians can be influenced to an undue extent by recent encounters with selected patients. Evidence-based medicine has become an increasingly important part of routine medical practice and has led to the publication of many useful practice guidelines. It is important to remember, however, that only a small fraction of the many decisions made in clinical practice are based on rigorous clinical trial evidence; other guideline recommendations are, therefore, predicated on expert consensus and weaker evidentiary support.

Thus, the importance of evidence-based medicine notwithstanding, much medical decision-making still relies on good clinical judgment, an attribute that is difficult to quantify or even to assess qualitatively. Physicians must use their knowledge and experience as a basis for weighing known factors, along with the inevitable uncertainties, and then making a sound judgment; this synthesis of information is particularly important when a relevant evidence base is not available. Several quantitative tools may be invaluable in synthesizing the available information, including diagnostic tests, Bayes' theorem (the probability of an event predicated on prior knowledge of conditions possibly related to the event), and multivariate statistical models **(Chap. 4)**. Diagnostic tests serve to reduce uncertainty about an individual's diagnosis or prognosis and help the physician decide how best to manage that individual's condition. The battery of diagnostic tests complements the history and physical examination. The accuracy of a particular test is ascertained by determining its sensitivity (true-positive rate) and specificity (true-negative rate), as well as the predictive value of a positive and a negative result. **See Chap. 4 for a more thorough discussion of decision-making in clinical medicine.**

Practice Guidelines

Many professional organizations and government agencies have developed formal clinical-practice guidelines to aid physicians and other caregivers in making diagnostic and therapeutic decisions that are evidence-based, cost-effective, and most appropriate to a particular patient and clinical situation. As the evidence base of medicine increases, guidelines can provide a useful framework for managing patients with particular diagnoses or symptoms. Clinical guidelines can protect patients—particularly those with inadequate health care benefits—from receiving substandard care. These guidelines also can protect conscientious caregivers from inappropriate charges of malpractice and society from the excessive costs associated with the overuse of medical resources. There are, however, caveats associated with clinical-practice guidelines since they tend to oversimplify the complexities of medicine. Furthermore, groups with different perspectives may develop divergent recommendations regarding issues as basic as the need for screening of women by mammography or of men with serum prostate-specific antigen (PSA) measurements. Finally, guidelines, as the term implies, do not—and cannot be expected to—account for the uniqueness of each individual and his or her illness. The physician's challenge is to integrate into clinical practice the useful recommendations offered by experts without accepting them blindly or being inappropriately constrained by them.

Precision Medicine

The concept of *precision* or *personalized medicine* reflects the growing recognition that diseases once lumped together can be further stratified on the basis of genetic, biomarker, phenotypic, and/or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations. Inherent in this concept is the goal of targeting therapies in a more specific way to improve clinical outcomes for the individual patient and minimize unnecessary side effects for those less likely to respond to a particular treatment. In some respects, precision medicine represents the evolution of clinical practice guidelines, which are usually developed for populations of patients or a particular diagnosis (e.g., hypertension, thyroid nodule). As the pathobiology, prognosis, and treatment responses of subgroups within these diagnoses become better understood (i.e., through refined genomic analysis or enhanced deep phenotyping), the relevant clinical guidelines incorporate progressively more refined recommendations for individuals within these subgroups. The role of precision medicine is best illustrated for cancers in which genetic testing is able to predict responses (or the lack thereof) to targeted therapies **(Chap. 73)**. One can anticipate similar applications of precision medicine in pharmacogenomics,

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Evaluation of Outcomes

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

and minimize unnecessary side effects for those less likely to respond to a particular treatment. In some respects, precision medicine represents the for individuals within these subgroups. The role of precision medicine is best illustrated for cancers in which genetic testing is able to predict responses (or the lack thereof) to targeted therapies (Chap. 73). One can anticipate similar applications of precision medicine in pharmacogenomics, immunologic disorders, and diseases in which biomarkers can predict treatment responses. **See Chap. 5 for a more thorough discussion of precision medicine.**

Evaluation of Outcomes

Clinicians generally use *objective* and readily measurable parameters to judge the outcome of a therapeutic intervention. These measures may oversimplify the complexity of a clinical condition as patients often present with a major clinical problem in the context of multiple complicating background illnesses. For example, a patient may present with chest pain and cardiac ischemia, but with a background of chronic obstructive pulmonary disease and renal insufficiency. For this reason, outcome measures, such as mortality, length of hospital stay, or readmission rates, are typically risk-adjusted. An important point to remember is that patients usually seek medical attention for *subjective* reasons; they wish to obtain relief from pain, to preserve or regain function, and to enjoy life. The components of a patient's health status or quality of life can include bodily comfort, capacity for physical activity, personal and professional function, sexual function, cognitive function, and overall perception of health. Each of these important domains can be assessed through structured interviews or specially designed questionnaires. Such assessments provide useful parameters by which a physician can judge patients' subjective views of their disabilities and responses to treatment, particularly in chronic illness. The practice of medicine requires consideration and integration of both objective and subjective outcomes.

Many health systems use survey and patient feedback data to assess qualitative features such as patient satisfaction, access to care, and communication with nurses and physicians. In the United States, HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) surveys are used by many systems and are publicly reported. Social media is also being used to assess feedback in real time as well as to share patient experiences with health care systems, potentially enriching the information available for use in medical decisions.

Errors in the Delivery of Health Care

A series of reports from the Institute of Medicine (now the National Academy of Medicine [NAM]) called for an ambitious agenda to reduce medical error rates and improve patient safety by designing and implementing fundamental changes in health care systems (Chap. 8). It is the responsibility of hospitals and health care organizations to develop systems to reduce risk and ensure patient safety. Medication errors can be reduced through the use of ordering systems that rely on electronic processes or, when electronic options are not available, that eliminate misreading of handwriting. Whatever the clinical situation, it is the physician's responsibility to use powerful therapeutic measures wisely, with due regard for their beneficial actions, potential dangers, and cost. Implementation of infection control systems, enforcement of hand-washing protocols, and careful oversight of antibiotic use can minimize the complications of nosocomial infections. Central-line infection rates and catheter-associated urinary tract infections have been dramatically reduced at many centers by careful adherence of trained personnel to standardized protocols for introducing and maintaining central lines and urinary catheters, respectively. Rates of surgical infection and wrong-site surgery can likewise be reduced by the use of standardized protocols and checklists. Falls by patients can be minimized by judicious use of sedatives and appropriate assistance with bed-to-chair and bed-to-bathroom transitions. Taken together, these and other measures are saving thousands of lives each year.

Electronic Medical Records

Both the growing reliance on computers and the strength of information technology now play central roles in medicine, including efforts to reduce medical errors. Laboratory data are accessed almost universally through computers. Many medical centers now have electronic medical records (EMRs), computerized order entry, and bar-coded tracking of medications. Some of these systems are interactive, sending reminders or warning of potential medical errors.

EMRs offer rapid access to information that is invaluable in enhancing health care quality and patient safety, including relevant data, historical and clinical information, imaging studies, laboratory results, and medication records. These data can be used to monitor and reduce unnecessary variations in care and to provide real-time information about processes of care and clinical outcomes. Ideally, patient records are easily transferred across the health care system; however, technological limitations and concerns about privacy and cost continue to limit broad-based use of EMRs in many clinical settings.

For all of the advantages of EMRs, they can create distance between the physician and patient if care is not taken to preserve face-to-face contact. EMRs also require training and time for data entry. Many providers spend significant time entering information to generate structured data and to meet billing requirements. They may feel pressured to take short cuts, such as "cutting and pasting" parts of earlier notes into the daily record, thereby

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

of dissatisfaction among physicians, lessening professional and personal well-being as a result. Clearly, this is an area of daily practice that requires improvement both for the delivery of safe and optimal care and physician wellness.

across the health care system; however, technological limitations and concerns about privacy and cost continue to limit broad use of EMRs in also require training and time for data entry. Many providers spend significant time entering information to generate structured data and to meet billing requirements. They may feel pressured to take short cuts, such as "cutting and pasting" parts of earlier notes into the daily record, thereby increasing the risk of errors. EMRs also structure information in a manner that disrupts the traditional narrative flow across time and among providers. These features, which may be frustrating for some providers, must be weighed against the advantages of ready access to past medical history, imaging, laboratory data, and consultant notes. Furthermore, the effort, time, and attention needed to maintain and utilize the EMR have led to a growing sense of dissatisfaction among physicians, lessening professional and personal well-being as a result. Clearly, this is an area of daily practice that requires improvement both for the delivery of safe and optimal care and physician wellness.

It is important to emphasize that information technology is merely a tool and can never replace the clinical decisions that are best made by the physician. Clinical knowledge and an understanding of a patient's needs, supplemented by quantitative tools, still represent the best approach to decision-making in the practice of medicine.

THE PATIENT–PHYSICIAN RELATIONSHIP

> *The significance of the intimate personal relationship between physician and patient cannot be too strongly emphasized, for in an extraordinarily large number of cases both the diagnosis and treatment are directly dependent on it. One of the essential qualities of the clinician is interest in humanity,* ***for the secret of the care of the patient is in caring for the patient****.*
>
> —Francis W. Peabody, October 21, 1925, Lecture at Harvard Medical School

Physicians must never forget that patients are individuals with problems that all too often transcend their physical complaints. They are not "cases" or "admissions" or "diseases." Patients do not fail treatments; treatments fail to benefit patients. This point is particularly important in this era of high technology in clinical medicine. Most patients are anxious and fearful. Physicians should instill confidence and offer reassurance, but they must never come across as arrogant, patronizing, impatient, or hurried. A professional attitude, coupled with warmth and openness, can do much to alleviate anxiety and to encourage patients to share all aspects of their medical history. Empathy and compassion are the essential features of a caring physician. The physician needs to consider the setting in which an illness occurs—in terms not only of patients themselves but also of their familial, social, and cultural backgrounds. The ideal patient–physician relationship is based on thorough knowledge of the patient, mutual trust, and the ability to communicate.

Informed Consent

The fundamental principles of medical ethics require physicians to act in the patient's best interest and to respect the patient's autonomy. Both principles are reflected in the process of informed consent. Patients are required to sign consent forms for most diagnostic or therapeutic procedures. Many patients possess limited medical knowledge and must rely on their physicians for advice. Communicating in a clear and understandable manner, physicians must fully discuss the alternatives for care and explain the risks, benefits, and likely consequences of each alternative. The physician is responsible for ensuring that the patient thoroughly understands these risks and benefits; encouraging questions is an important part of this process. It may be necessary to go over certain issues with the patient more than once. This is the very definition of *informed consent*. Complete, clear explanation and discussion of the proposed procedures and treatment can greatly mitigate the fear of the unknown that commonly accompanies hospitalization. Often the patient's understanding is enhanced by repeatedly discussing the issues in an unthreatening and supportive way, answering new questions that occur to the patient as they arise. Continuing efforts to educate the patient are essential. Patients are frequently inhibited from understanding by the fear of an uncertain future and potential impact of the illness on themselves and their families. Clear communication can also help alleviate misunderstandings in situations where complications of intervention occur. Special care should also be taken to ensure that a physician seeking a patient's informed consent has no real or apparent conflict of interest.

Approach to Grave Prognoses and Death

No circumstance is more distressing than the diagnosis of an incurable disease, particularly when premature death is inevitable. What should the patient and family be told? What measures should be taken to maintain life? What can be done to optimize quality of life?

Transparency of information, delivered in an appropriate manner, is essential in the face of a terminal illness. Even patients who seem unaware of their medical circumstances, or whose family members have protected them from diagnoses or prognoses, often have keen insights into their condition. They may also have misunderstandings that can lead to additional anxiety. The patient must be given an opportunity to speak with the physician and ask questions. A wise and insightful physician uses such open communication as the basis for assessing what the patient wants to know

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

beliefs should be taken into consideration. Some patients may find it easier to share their feelings about death with their physician, nurses, or members of the clergy than with family members or friends.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

condition. They may also have misunderstandings that can lead to additional anxiety. The patient must be given an opportunity to speak with the physician and ask questions. A wise and insightful physician uses such open communication as the basis for assessing what the patient wants to know and when he or she wants to know it. On the basis of the patient's responses, the physician can assess the most appropriate time and pace for sharing information. Ultimately, the patient must understand the expected course of the disease so that appropriate plans and preparations can be made. The patient should participate in decision-making with an understanding of the goal of treatment (palliation) and its likely effects. The patient's religious beliefs should be taken into consideration. Some patients may find it easier to share their feelings about death with their physician, nurses, or members of the clergy than with family members or friends.

The physician should provide or arrange for emotional, physical, and spiritual support, and must be compassionate, unhurried, and open. In many instances, there is much to be gained by the laying on of hands. Pain should be controlled adequately, human dignity maintained, and isolation from family and close friends avoided. These aspects of care tend to be overlooked in hospitals, where the intrusion of life-sustaining equipment can detract from attention to the individual person and encourage concentration instead on the life-threatening disease, against which the battle ultimately will be lost in any case. In the face of terminal illness, the goal of medicine must shift from *cure* to *care* in the broadest sense of the term. *Primum succurrere*, first to help, is a guiding principle. In offering care to a dying patient, a physician should be prepared to provide information to family members and deal with their grief and sometimes their feelings of guilt or even anger. It is important for the physician to assure the family that everything reasonable is being done. A substantial challenge in these discussions is that the physician often does not know exactly how to gauge the prognosis. In addition, various members of the health care team may offer different opinions. Good communication among providers is essential so that consistent information is provided to patients. This is especially important when the best path forward is uncertain. Advice from experts in palliative and terminal care should be sought whenever appropriate to ensure that clinicians are not providing patients with unrealistic expectations.
For a more complete discussion of end-of-life care, see Chap. 12.

Maintaining Humanism and Professionalism

Many trends in the delivery of health care tend to make medical care impersonal. These trends, some of which have been mentioned already, include (1) vigorous efforts to reduce the escalating costs of health care; (2) the growing number of managed-care programs, which are intended to reduce costs but where the patient may have little choice in selecting a physician; (3) increasing reliance on technological advances and computerization; and (4) the need for numerous physicians and other health professionals to be involved in the care of most patients who are seriously ill.

In light of these changes in the medical care system, it is a major challenge for physicians to maintain the *humane* aspects of medical care. The American Board of Internal Medicine, working together with the American College of Physicians–American Society of Internal Medicine and the European Federation of Internal Medicine, has published a *Charter on Medical Professionalism* that underscores three main principles in physicians' contract with society: (1) the primacy of patient welfare, (2) patient autonomy, and (3) social justice. While medical schools appropriately place substantial emphasis on professionalism, a physician's personal attributes, including integrity, respect, and compassion, also are extremely important. In the United States, the Gold Humanism Society recognizes individuals who are exemplars of humanistic patient care and serve as role models for medical education and training.

Availability to the patient, expression of sincere concern, willingness to take the time to explain all aspects of the illness, and a nonjudgmental attitude when dealing with patients whose cultures, lifestyles, attitudes, and values differ from those of the physician are just a few of the characteristics of a humane physician. Every physician will, at times, be challenged by patients who evoke strongly negative or positive emotional responses. Physicians should be alert to their own reactions to such situations and should consciously monitor and control their behavior so that the patient's best interest remains the principal motivation for their actions at all times.

Another important aspect of patient care involves an appreciation of the patient's "quality of life," a subjective assessment of what each patient values most. This assessment requires detailed, sometimes intimate knowledge of the patient, which usually can be obtained only through deliberate, unhurried, and often repeated conversations. Time pressures will always threaten these interactions, but they should not diminish the importance of understanding and seeking to fulfill the priorities of the patient.

EXPANDING FRONTIERS IN MEDICAL PRACTICE

The Era of "Omics"

In the spring of 2003, announcement of the complete sequencing of the human genome officially ushered in the genomic era. However, even before that landmark accomplishment, the practice of medicine had been evolving as a result of insights into both the human genome and the genomes of a

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

assess the risk of certain diseases as well as variations in response to a number of drugs. Despite these advances, the use of complex genomics in the diagnosis, prevention, and treatment of disease is still in its early stages. The task of physicians is complicated by the fact that phenotypes generally are

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

In the spring of 2003, announcement of the complete sequencing of the human genome officially ushered in the genomic era. However, even before that landmark accomplishment, the practice of medicine had been evolving as a result of insights into both the human genome and the genomes of a wide variety of microbes. The clinical implications of these insights are illustrated by the complete genome sequencing of H1N1 influenza virus in 2009 and even faster sequencing of COVID-19 in early 2020, leading to the swift development and dissemination of effective vaccines. Today, gene expression profiles are being used to guide therapy and inform prognosis for a number of diseases, and genotyping is providing a new means to assess the risk of certain diseases as well as variations in response to a number of drugs. Despite these advances, the use of complex genomics in the diagnosis, prevention, and treatment of disease is still in its early stages. The task of physicians is complicated by the fact that phenotypes generally are determined not by genes alone but by the complex interactions among genes and gene products, and by the interplay of genetic and environmental factors.

Rapid progress is also being made in other areas of molecular medicine. *Epigenetics* is the study of alterations in chromatin and histone proteins and methylation of DNA sequences that influence gene expression **(Chap. 483)**. Every cell of the body has identical DNA sequences; the diverse phenotypes a person's cells manifest are, in part, the result of epigenetic regulation of gene expression. Epigenetic alterations are associated with a number of cancers and other diseases. *Proteomics*, the study of the entire library of proteins made in a cell or organ and the complex relationship of these proteins to disease, is enhancing the repertoire of the 23,000 genes in the human genome through alternate splicing, posttranslational processing, and posttranslational modifications that often have unique functional consequences. The presence or absence of particular proteins in the circulation or in cells is being explored for many diagnostic and disease-screening applications. *Microbiomics* is the study of the resident microbes in humans and other mammals, which together compose the microbiome. The human haploid genome has ~23,000 genes, whereas the microbes residing on and in the human body encompass more than 3–4 million genes; these resident microbes are likely to be of great significance with regard to health status. Ongoing research is demonstrating that the microbes inhabiting human mucosal and skin surfaces play a critical role in maturation of the immune system, in metabolic balance, in brain function, and in disease susceptibility. A variety of environmental factors, including the use and overuse of antibiotics, have been tied experimentally to substantial increases in disorders such as obesity, metabolic syndrome, atherosclerosis, and immune-mediated diseases in both adults and children. *Metagenomics*, of which microbiomics is a part, is the genomic study of environmental species that have the potential to influence human biology directly or indirectly. An example is the study of exposures to microorganisms in farm environments that may be responsible for the lower incidence of asthma among children raised on farms. *Metabolomics* is the study of the range of metabolites in cells or organs and the ways they are altered in disease states. The aging process itself may leave telltale metabolic footprints that allow the prediction (and possibly the prevention) of organ dysfunction and disease. It seems likely that disease-associated patterns will be found in lipids, carbohydrates, membranes, mitochondria and mitochondrial function, and other vital components of cells and tissues. *Exposomics* is the study of the exposome—i.e., the environmental exposures such as smoking, sunlight, diet, exercise, education, and violence that together have an enormous impact on health. All of this new information represents a challenge to the traditional reductionist approach to medical thinking. The variability of results in different patients, together with the large number of variables that can be assessed, creates challenges in identifying preclinical disease and defining disease states unequivocally. Accordingly, the tools of *systems biology* and *network medicine* are being applied to the enormous body of information ("big data") now obtainable for every patient and may eventually provide new approaches to classifying disease. **For a more complete discussion of a complex systems and network science approach to human disease, see Chap. 486.**

The rapidity of these advances may seem overwhelming to practicing physicians; however, physicians have an important role to play in ensuring that these powerful technologies and sources of new information are applied judiciously to patient care. Since omics are evolving so rapidly, physicians and other health care professionals must engage in continuous learning so that they can apply this new knowledge to the benefit of their patients' health and well-being. Genetic testing requires wise counsel based on an understanding of the value and limitations of the tests as well as the implications of their results for specific individuals. **For a more complete discussion of genetic testing, see Chap. 467.**

The Globalization of Medicine

Physicians should be cognizant of diseases and health care services beyond local boundaries. Global travel has critical implications for disease spread, and it is not uncommon for diseases endemic to certain regions to be seen in other regions after a patient has traveled to and returned from those regions. The outbreak of Zika virus infections in the Americas is a cogent example of this phenomenon. In addition, factors such as wars, the migration of refugees, and increasing climate extremes are contributing to changing disease profiles worldwide. Patients have broader access to unique expertise or clinical trials at distant medical centers, even those in other countries, and the cost of travel may be offset by the quality of care at those distant locations. As much as any other factor influencing global aspects of medicine, the Internet has transformed the transfer of medical information throughout the world. This change has been accompanied by the transfer of technological skills through telemedicine and international consultation —for example, interpretation of radiologic images and pathologic specimens. **For a complete discussion of global issues, see Chap. 472.**

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

to physicians and patients almost instantaneously at any time and from anywhere in the world. This medium holds enormous potential for the delivery of current information, practice guidelines, state-of-the-art conferences, journal content, textbooks (including this text), and direct communications

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

—for example, interpretation of radiologic images and pathologic specimens. **For a complete discussion of global issues, see Chap. 472.**

Medicine on the Internet

On the whole, the Internet has had a positive effect on the practice of medicine; through personal computers, a wide range of information is available to physicians and patients almost instantaneously at any time and from anywhere in the world. This medium holds enormous potential for the delivery of current information, practice guidelines, state-of-the-art conferences, journal content, textbooks (including this text), and direct communications with other physicians and specialists, expanding the depth and breadth of information available to the physician regarding the diagnosis and care of patients. Medical journals are now accessible online, providing rapid sources of new information. By bringing them into direct and timely contact with the latest developments in medical care, this medium also serves to lessen the information gap that has hampered physicians and health care providers in remote areas.

Patients, too, are turning to the Internet in increasing numbers to acquire information about their illnesses and therapies and to join Internet-based support groups. Patients often arrive at a clinic visit with sophisticated information about their illnesses. In this regard, physicians are challenged in a positive way to keep abreast of the latest relevant information while serving as an "editor" as patients navigate this seemingly endless source of information, the accuracy and validity of which are not uniform.

A critically important caveat is that virtually anything can be published on the Internet, with easy circumvention of the peer-review process that is an essential feature of academic publications. Both physicians and patients who search the Internet for medical information must be aware of this danger. Notwithstanding this limitation, appropriate use of the Internet is revolutionizing information access for physicians and patients, and in this regard represents a remarkable resource that was not available to practitioners a generation ago.

Public Expectations and Accountability

The general public's level of knowledge and sophistication regarding health issues has grown rapidly over the past few decades. As a result, expectations of the health care system in general and of physicians in particular have risen. Physicians are expected to master rapidly advancing fields (the *science* of medicine) while considering their patients' unique needs (the *art* of medicine). Thus, physicians are held accountable not only for the technical aspects of the care they provide but also for their patients' satisfaction with the delivery and costs of care.

In many parts of the world, physicians increasingly are expected to account for the way in which they practice medicine by meeting certain standards prescribed by federal and local governments. The hospitalization of patients whose health care costs are reimbursed by the government and other third parties is subjected to utilization review. Thus, a physician must defend the cause for and duration of a patient's hospitalization if it falls outside certain "average" standards. Authorization for reimbursement increasingly is based on documentation of the nature and complexity of an illness, as reflected by recorded elements of the history and physical examination. A growing "pay-for-performance" movement seeks to link reimbursement to quality of care. The goal of this movement is to improve standards of health care and contain spiraling health care costs. In many parts of the United States, managed (capitated) care contracts with insurers have replaced traditional fee-for-service care, placing the onus of managing the cost of all care directly on the providers and increasing the emphasis on preventive strategies. In addition, physicians are expected to give evidence of their current competence through mandatory continuing education, patient record audits, maintenance of certification, and relicensing.

Medical Ethics and New Technologies

The rapid pace of technological advances has profound implications for medical applications that go far beyond the traditional goals of disease prevention, treatment, and cure. Cloning, genetic engineering, gene therapy, human–computer interfaces, nanotechnology, and use of targeted therapies have the potential to modify inherited predispositions to disease, select desired characteristics in embryos, augment "normal" human performance, replace failing tissues, and substantially prolong life span. Given their unique training, physicians have a responsibility to help shape the debate on the appropriate uses of and limits placed on these new technologies and to consider carefully the ethical issues associated with the implementation of such interventions. As medicine becomes more complex, shared decision-making is increasingly important, not only in areas such as genetic counseling and end-of-life care, but also in diagnostic and treatment options.

Learning Medicine

More than a century has passed since the publication of the Flexner Report, a seminal study that transformed medical education and emphasized the scientific foundations of medicine as well as the acquisition of clinical skills. In an era of burgeoning information and access to medical simulation and informatics, many schools are implementing new curricula that emphasize lifelong learning and the acquisition of competencies in teamwork,

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

the foundation for many of the themes highlighted in this chapter and are expected to allow physicians to progress, with experience and learning over time, from competency to proficiency to mastery.

scientific foundations of medicine as well as the acquisition of clinical skills. In an era of burgeoning information and access to medical simulation and informatics, many schools are implementing new curricula that emphasize lifelong learning and the acquisition of competencies in teamwork, communication skills, system-based practice, and professionalism. The tools of medicine also change continuously, necessitating formal training in the use of EMRs, large datasets, ultrasound, robotics, and new imaging techniques. These and other features of the medical school curriculum provide the foundation for many of the themes highlighted in this chapter and are expected to allow physicians to progress, with experience and learning over time, from competency to proficiency to mastery.

At a time when the amount of information that must be mastered to practice medicine continues to expand, increasing pressures both within and outside of medicine have led to the implementation of restrictions on the amount of time a physician-in-training can spend in the hospital and in clinics. Because the benefits associated with continuity of medical care and observation of a patient's progress over time were thought to be outstripped by the stresses imposed on trainees by long hours and by fatigue-related errors, strict limits were set on the number of patients that trainees could be responsible for at one time, the number of new patients they could evaluate in a day on call, and the number of hours they could spend in the hospital. In 1980, residents in medicine worked in the hospital more than 90 hours per week on average. In 1989, their hours were restricted to no more than 80 per week. Resident physicians' hours further decreased by ~10% between 1996 and 2008, and in 2010, the Accreditation Council for Graduate Medical Education further restricted (i.e., to 16 hours per shift) consecutive in-hospital duty hours for first-year residents. The impact of these changes is still being assessed, but the evidence that medical errors have decreased as a consequence is sparse. An unavoidable by-product of fewer hours at the bedside is an increase in the number of "handoffs" of patient responsibility from one physician to another. These transfers often involve a transition from a physician who knows the patient well, having evaluated that individual on admission, to a physician who knows the patient less well. It is imperative that these transitions of responsibility be handled with care and thoroughness, with all relevant information exchanged and acknowledged. These issues highlight the challenge our profession has in establishing a reliable measure of physician effectiveness.

The Physician as Perpetual Student

From the time physicians graduate from medical school, it becomes all too apparent that this milestone is symbolic and that they must embrace the role of a "perpetual student." This realization is at the same time exhilarating and anxiety-provoking. It is exhilarating because physicians can apply constantly expanding knowledge to the treatment of their patients; it is anxiety-provoking because physicians realize that they will never know as much as they want or need to know. Ideally, physicians will translate the latter feeling into energy through which they can continue to improve and reach their potential. It is the physician's responsibility to pursue new knowledge continually by reading, attending conferences and courses, and consulting colleagues and the Internet. This is often a difficult task for a busy practitioner; however, a commitment to continued learning is an integral part of being a physician and must be given the highest priority.

The Physician as Citizen

Being a physician is a privilege. The capacity to apply one's skills for the benefit of fellow human beings is a noble calling. The physician–patient relationship is inherently unbalanced in the distribution of power. In light of their influence, physicians must always be aware of the potential impact of what they do and say, and must always strive to strip away individual biases and preferences to find what is best for their patients. To the extent possible, physicians should also act within their communities to promote health and alleviate suffering. Meeting these goals begins by setting a healthy example and continues in taking action to deliver needed care even when personal financial compensation may not be available.

Research, Teaching, and the Practice of Medicine

The word *doctor* is derived from the Latin *docere*, "to teach." As teachers, physicians should share information and medical knowledge with colleagues, students of medicine and related professions, and their patients. The practice of medicine is dependent on the sum total of medical knowledge, which in turn is based on an unending chain of scientific discovery, clinical observation, analysis, and interpretation. Advances in medicine depend on the acquisition of new information through research, and improved medical care requires the transmission of that information. As part of their broader societal responsibilities, physicians should encourage patients to participate in ethical and properly approved clinical investigations if these studies do not impose undue hazard, discomfort, or inconvenience. Physicians engaged in clinical research must be alert to potential conflicts of interest between their research goals and their obligations to individual patients. The best interests of the patient must always take priority.

> *To wrest from nature the secrets which have perplexed philosophers in all ages, to track to their sources the causes of disease, to correlate the vast stores of knowledge, that they may be quickly available for the prevention and cure of disease—these are our ambitions.*

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FURTHER READING

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

stores of knowledge, that they may be quickly available for the prevention and cure of disease—these are our ambitions.

—William Osler, 1849–1919

FURTHER READING

Cheston CC et al: Social media use in medical education: A systematic review. Acad Med 88:893, 2013. [PubMed: 23619071]

Cooke M et al: American medical education 100 years after the Flexner report. N Engl J Med 355:1339, 2006. [PubMed: 17005951]

Excel JL et al: Vaccine development for emerging infectious diseases. Nat Med 27:591, 2021 [PubMed: 33846611]

Institute of Medicine: *Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life*. Washington, DC, National Academies Press, 2015. [PubMed:] [[XSLOpenURL/]]

Institute of Medicine: *Improving Diagnosis in Health Care*. Washington, DC, National Academies of Sciences, Engineering, and Medicine, 2015.

Levine DM et al: Hospital-level care at home for acutely ill adults: A qualitative evaluation of a randomized controlled trial. J Gen Intern Med 36:1965, 2021.

Stern DT, Papadakis M: The developing physician—becoming a professional. N Engl J Med 355:1794, 2006. [PubMed: 17065641]

Vickrey BG et al: How neurologists think: A cognitive psychology perspective on missed diagnoses. Ann Neurol 67:425, 2010. [PubMed: 20437577]

West P et al: Intervention to promote physician well-being, job satisfaction, and professionalism. A randomized clinical trial. JAMA Intern Med 174:527, 2014. [PubMed: 24515493]

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 1: The Practice of Medicine, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

Chapter 2: Promoting Good Health

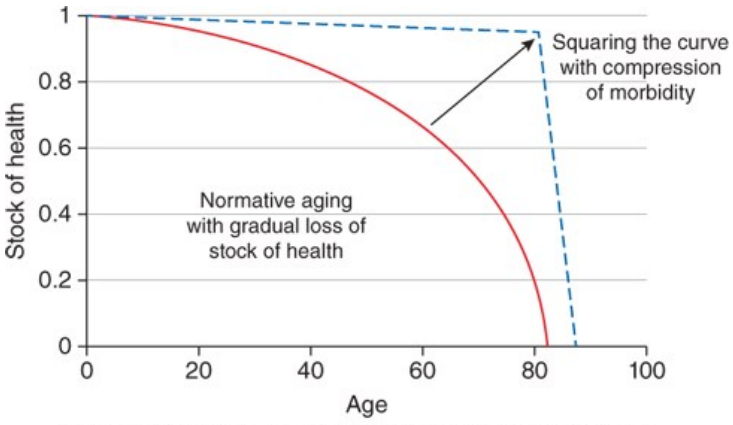
Donald M. Lloyd-Jones; Kathleen M. McKibbin

GOALS AND APPROACHES TO PREVENTION

Prevention of acute and chronic diseases before their onset has been recognized as one of the hallmarks of excellent medical practice for centuries and is now used as a metric for highly functioning health care systems. The ultimate goal of preventive strategies is to avoid premature death. However, as longevity has increased dramatically worldwide over the last century (largely as a result of public health practices), increasing emphasis is placed on prevention for the purpose of preserving quality of life and extending the health span, not just the life span. Given that all patients will eventually die, the goal of prevention ultimately becomes compression of morbidity toward the end of the life span; that is, reduction of the amount of burden and time spent with disease prior to dying. As shown in **Fig. 2-1**, normative aging tends to involve a steady decline in the stock of health, with accelerating decline over time. Successful prevention offers the opportunity both to extend life and to extend healthy life, thus "squaring the curve" of health loss during aging.

FIGURE 2-1

Loss of health with aging. Representation of normative aging with loss of the full stock of health with which individuals are born (indicating gain of morbidity), contrasted with a squared curve with greater longevity and fuller stock of health (less morbidity) until shortly before death. The "squared curve" represents the likely ideal situation for most patients.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Prevention strategies have been characterized as tertiary, secondary, primary, and primordial. *Tertiary prevention* requires rapid action to prevent imminent death in the setting of acute illness, such as through percutaneous coronary intervention in the setting of ST-segment elevation myocardial infarction. *Secondary prevention* strategies focus on avoiding the recurrence of disease and death in an individual who is already affected. For example, tamoxifen is recommended for women with surgically treated early-stage, estrogen receptor–positive breast cancer, because it reduces the risk of recurrent breast cancer (including in the contralateral breast) and death. *Primary prevention* attempts to reduce the risk of incident disease among individuals with one or more risk factors. Treatment of elevated blood pressure in individuals who have not yet experienced cardiovascular disease represents one example of primary prevention that has proven effective in reducing the incidence of stroke, heart failure, and coronary heart disease.

Primordial prevention is a more recent concept (first introduced in 1979) that focuses on prevention of the development of *risk factors* for disease, not just prevention of disease. Primordial prevention strategies emphasize upstream determinants of risk for chronic diseases, such as eating patterns, physical activity, and environmental and social determinants of health. It therefore encompasses medical treatment strategies for some individuals as developed the risk factor. The time spent with adverse levels of the risk factor often causes irreversible damage that precludes complete restoration of low risk. For example, individuals with hypertension who are treated back to optimal levels (<120/<80 mmHg) do have a lower risk compared with

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

disease represents one example of primary prevention that has proven effective in reducing the incidence of stroke, heart failure, and coronary heart

ACCESS Medicine

Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

just prevention of disease. Primordial prevention strategies emphasize upstream determinants of risk for chronic diseases, such as eating patterns, physical activity, and environmental and social determinants of health. It therefore encompasses medical treatment strategies for some individuals as well as a strong reliance on public health and social policy. It is increasingly clear that primordial prevention represents the ultimate means for reducing the burden of chronic diseases of aging. Once risk factors develop, it is difficult to restore risk to the low level of someone who never developed the risk factor. The time spent with adverse levels of the risk factor often causes irreversible damage that precludes complete restoration of low risk. For example, individuals with hypertension who are treated back to optimal levels (<120/<80 mmHg) do have a lower risk compared with untreated patients with hypertension, but they still have twice the risk of cardiovascular events as those who maintained optimal blood pressure without medications. Patients with elevated blood pressure that is subsequently treated have greater left ventricular mass index, worse renal function, and more evidence of atherosclerosis and other target organ damage as a result of the time spent with elevated blood pressure; such damage cannot be fully reversed despite efficacious therapy with antihypertensive medications. Conversely, as described below in greater detail, individuals who maintain optimal levels of all major cardiovascular risk factors into middle age through primordial prevention essentially abolish their lifetime risk of developing cardiovascular disease while also living substantially longer and having a lower burden and later onset of other comorbid illnesses (compression of morbidity).

Prevention strategies should be distinguished from disease screening strategies. Screening attempts to detect evidence of disease at its earliest stages, when treatment is likely to be more efficacious than for advanced disease **(Chap. 6)**. Screening can be performed in service of prevention, especially if it aids in identifying preclinical markers, such as dyslipidemia or hyperglycemia, associated with elevated disease risk.

HEALTH PROMOTION

In recent decades, medical practice has increasingly focused on clinical and public health approaches to promote health, and not just prevent disease. Prevention of disease is a worthy individual and societal goal in and of itself, but it does not necessarily guarantee health. Health is a broader construct encompassing more than just absence of disease. It includes biologic, physiologic, and psychological domains (among others) in a continuum, rather than occurring as a dichotomous trait. Health is therefore somewhat subjective, but attempts have been made to use more objective criteria to define health in order to raise awareness, prevent disease, and promote healthy longevity.

For example, in 2010, the American Heart Association (AHA) defined a new construct of “cardiovascular health” based on evidence of associations with longevity, disease avoidance, healthy longevity, and quality of life. The definition of cardiovascular health is based on seven health behaviors and health factors (eating pattern, physical activity, smoking status, body mass index [BMI], and levels of blood pressure, blood cholesterol, and blood glucose) and includes a spectrum from poor to ideal. Individuals with optimal levels of all seven metrics simultaneously are considered to have ideal cardiovascular health. The state of cardiovascular health for an individual or a population can be assessed with simple scoring by counting the number of ideal metrics (out of seven) or applying 0 points for each poor metric, 1 point for each intermediate metric, and 2 points for each ideal metric, thus creating a composite cardiovascular health score ranging from 0 to 14 points. Higher cardiovascular health scores in younger and middle ages have been associated with greater longevity, lower incidence of cardiovascular disease, lower incidence of other chronic diseases of aging (including dementia, cancer, and more), compression of morbidity, greater quality of life, and lower health care costs, achieving both individual and societal goals for healthy aging and further establishing the critical importance of primordial prevention and cardiovascular health promotion.

Focusing on health promotion, rather than just disease prevention, may also provide greater motivation for patients to pursue lifestyle changes or adhere to clinician recommendations. Extensive literature suggests that providing patients solely with information regarding disease risk, or risk reduction with treatment, is unlikely to motivate desired behavior change. Empowering patients with strategies to achieve positive health goals after discussing risks can provide more effective adherence and better long-term outcomes. In the case of smoking cessation, enumerating only the risks of smoking can lead to patient inertia and therapeutic nihilism and has proven to be an ineffective approach, whereas strategies that incorporate positive health messaging, support, and feedback, with appropriate use of evidence-based therapies, have proven far more effective.

PRIORITIZING PREVENTION STRATEGIES

In secondary prevention, the patient already has manifest clinical disease and is therefore at high risk for progression. The approach should be to work with the patient to implement all evidence-based strategies that will help to prevent recurrence or progression. This will typically include drug therapy as well as therapeutic lifestyle changes to control ongoing risk factors that may have caused disease in the first place. Juggling priorities can be difficult, and barriers to implementation are many, including costs, time, patient health literacy, and patient and caregiver capacity to organize the regimen. Addressing these potential barriers with the patient can help to forge a therapeutic bond and may improve adherence; ignoring them will likely lead to therapeutic failure. Numerous studies demonstrate that, even in high-functioning health systems, only ~50% of patients are taking

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

individual patient’s likelihood of benefit from a given preventive intervention. A first step is to understand the patient’s estimated absolute risk for disease in the foreseeable future or during their remaining life span. However, absolute risk estimation and presentation of those risks are generally

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

regimen. Addressing these potential barriers with the patient can help to forge a therapeutic bond and may improve adherence; ignoring them will likely lead to therapeutic failure. Numerous studies demonstrate that, even in high-functioning health systems, only ~50% of patients are taking recommended, evidence-based secondary prevention medications, such as statins, by 1 year after a myocardial infarction.

In patients who are eligible for primary prevention strategies, it is important to frame the discussion around the overall evidence base as well as an individual patient's likelihood of benefit from a given preventive intervention. A first step is to understand the patient's estimated absolute risk for disease in the foreseeable future or during their remaining life span. However, absolute risk estimation and presentation of those risks are generally insufficient to motivate behavior change. It is critical to assess the patient's understanding and tolerance of the risk, their readiness to implement lifestyle changes or adhere to drug therapy, and their overall preferences regarding use of drug therapy to prevent an event (e.g., cancer, myocardial infarction, stroke). The clinician can help the patient by informing them of the risks for disease and potential for absolute benefits (and harms) from the available evidence-based choices. This may take more than one conversation, but given that diseases, such as cancer and cardiovascular disease, are the leading causes of premature death and disability, the time is well spent.

Partnering with the patient through motivational interviewing may assist in the process of selecting initial approaches to prevention. Selecting an area that the patient feels they are ready to change can lead to better adherence and greater achievement of success in the short and longer term. If the patient is uncertain what course to choose, prudence would dictate focusing on control of risk factors that may lead to the most rapid reduction in risk for acute events. For example, blood pressure is both a chronic risk factor and an acute trigger for cardiovascular events. Thus, if a patient has both significant elevations in blood pressure and dyslipidemia, it would be appropriate to focus initial efforts on blood pressure control. Likewise, a focus on smoking cessation can lead to more rapid reductions in risk for acute events than some other lifestyle interventions.

PREVENTION AND HEALTH PROMOTION ACROSS THE LIFE COURSE

Periodic Health Evaluations

The "routine annual physical" has in many ways become an expected part of the patient-physician relationship in primary care practice. However, evidence for the efficacy of the periodic health evaluation in asymptomatic adults unselected for risk factors or disease is mixed and depends on the outcome. Systematic reviews and meta-analyses of published trials have consistently observed lack of benefit (and also lack of harm) in terms of total mortality in association with periodic health evaluations. Data are more heterogeneous but overall suggest no benefit for cancer- or cardiovascular-specific mortality, with the potential for either benefit or harm depending on number of evaluations and patient-level factors. Well-designed studies on nonfatal clinical events and morbidity have been sparsely reported, but there appear to be no large effects.

Periodic health evaluations do appear to lead to greater diagnosis of certain conditions such as hypertension and dyslipidemia, as expected. Likewise, periodic health examinations also improve the delivery of recommended preventive services, such as gynecologic examinations and Papanicolaou smears, fecal occult blood testing, and cholesterol screening. The benefits and risks associated with screening tests are discussed in detail in **Chap. 6**. Risks of routine evaluations include inappropriate testing or overtesting or false-positive findings that require follow-up and induce patients to worry. Periodic health examinations appear to be associated with less patient worry. On balance, given the lack of convincing evidence of harm and the potential for better delivery of appropriate screening, counseling, and preventive services, periodic health evaluations appear reasonable for general populations at average risk for chronic conditions.

It is important to note that routine annual comprehensive physical examinations of asymptomatic adult patients have very low yield and may take an inordinate amount of time in a wellness visit. Such time may be better spent on assessing and counseling the patient on other aspects of their health, as discussed below. Evidence-based components that should be included in periodic evaluations focused on health and prevention include a number of age-appropriate screening tests for chronic disease and risk factors, preventive interventions including immunizations and chemoprevention for at-risk individuals, and preventive counseling. The U.S. Preventive Services Task Force publishes its *Guide to Clinical Preventive Services,* which contains evidence-based recommendations from the Task Force on preventive services for which there is a high degree of certainty that the service provides at least moderate net clinical benefit (i.e., benefits outweigh harms significantly and to a reasonable magnitude).

Healthy Behaviors and Lifestyles

Owing to the paucity of evidence, the heterogeneity of study designs, and the diverse nature of interventions studied, many clinicians are uncertain as to how to deliver advice regarding healthy behaviors and lifestyles. Nevertheless, adverse behaviors and lifestyles contribute to >75% of premature, preventable deaths and disability. Estimates from the U.S. National Health and Nutrition Examination Survey indicate that fewer than 1% of Americans achieve an optimal heart-healthy eating pattern. Thus, whereas there are many demands on time during a typical patient-clinician encounter, few

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Healthy Eating Patterns

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

preventable deaths and disability. Estimates from the U.S. National Health and Nutrition Examination Survey indicate that fewer than 1% of Americans achieve an optimal heart-healthy eating pattern. Thus, whereas there are many demands on time during a typical patient-clinician encounter, few things may have more impact on longevity, health, and quality of life for asymptomatic patients than an efficient approach to assessing, documenting, and improving patients' health behaviors. Indeed, the mere act of assessing health behaviors has been shown to affect patients' health behaviors. Facility with tools for assessment of lifestyle and with strategies for counseling are therefore of paramount importance.

Healthy Eating Patterns

Despite the existence of numerous "fad" diets and seemingly inconsistent recommendations on dietary composition, there is remarkable agreement about what should constitute a healthy eating pattern for the broad population to avoid nutritional deficits (i.e., vitamin deficiency) and excesses (i.e., excessive caloric intake) and to maximize potential health **(see Chap. 332) (Table 2-1)**. Optimal eating patterns consist of whole fruits and vegetables, whole grains, lean proteins, and healthy oils, and allow for nonfat or low-fat dairy intake. They tend to exclude frequent ingestion of foods high in refined sugars and starches, saturated fat, and sodium. Since sodium and refined sugars and starches are the hallmark of much of the processed/packaged food supply, a simple rule of thumb is to provide or cook the majority of one's own meals starting from whole foods and emphasizing fruits and vegetables. Likewise, foods prepared outside of the home tend to have higher fat and sodium content, so special attention to menu choices focused on fruits, vegetables, lean proteins, and whole grains, while minimizing sauces and dressings, can help most individuals follow healthier eating patterns when eating food prepared outside the home. In all cases, sugar-sweetened beverages and nonnutritious snack foods should be minimized. If snacks are included, small amounts of healthy nuts and seeds or more fruits and vegetables should be encouraged.

TABLE 2-1

Guidelines and Key Recommendations from the *Dietary Guidelines for Americans*, 2020–2025

Guidelines	Key Recommendations
1. Follow a healthy dietary pattern at every life stage. For the first 6 months of life, infants should exclusively be fed human milk, or iron-fortified formula if human milk is unavailable. From 6 to 12 months, infants should be introduced to a variety of complementary nutrient-dense foods. From 12 months to older adulthood, the dietary pattern should meet nutrient needs, help achieve a healthy body weight, and reduce the risk of chronic disease. **2. Customize and enjoy nutrient-dense food and beverage choices to reflect personal preferences, cultural traditions, and budgetary considerations.** The Dietary Guidelines provide a framework of several dietary patterns intended to be customized to individual needs and preferences, as well as the foodways of the diverse cultures in the United States. **3. Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits.** Nutrient-dense foods provide vitamins, minerals, and other health-promoting components and have no or little added sugars, saturated fat, and sodium. A healthy dietary pattern consists of nutrient-dense forms of foods and beverages across all food groups, in recommended amounts, and within calorie limits. **4. Limit foods and beverages higher in added sugars, saturated fat, and sodium, and limit alcoholic beverages.** At every life stage, meeting food group recommendations, even with nutrient-dense choices, fulfills most of a person's daily calorie needs and sodium limits, with little room for extra added sugars, saturated fat, or sodium, or for alcoholic beverages.	**The Dietary Guidelines' Key Recommendations for healthy eating patterns should be applied in their entirety, given the interconnected relationship that each dietary component can have with others. They are also intended as a framework to accommodate personal preferences, cultural traditions, and budgetary considerations.** **Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits to achieve a healthy weight and reduce the risk of chronic disease.** **The core elements that make up a healthy dietary pattern include:** • Vegetables of all types—dark green; red and orange; beans, peas, and lentils; starchy; and other vegetables • Fruits, especially whole fruit • Grains, at least half of which are whole grain • Dairy, including fat-free or low-fat milk, yogurt, and cheese, and/or lactose-free versions and fortified soy beverages and yogurt as alternatives • Protein foods, including lean meats, poultry, and eggs; seafood; beans, peas, and lentils; and nuts, seeds, and soy products • Oils, including vegetable oils and oils in food, such as seafood and nuts **A healthy eating pattern limits:** • Added sugars—Less than 10% of calories per day starting at age 2. Avoid foods and beverages with added sugars for those younger than age 2. • Alcoholic beverages—Adults of legal drinking age can choose not to drink or to drink in moderation by limiting intake to 2 drinks or

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Guidelines	Key Recommendations
1. Follow a healthy dietary pattern at every life stage. For the first 6 months of life, infants should exclusively be fed human milk, or iron-fortified formula if human milk is unavailable. From 6 to 12 months, infants should be introduced to a variety of complementary nutrient-dense foods. From 12 months to older adulthood, the dietary pattern should meet nutrient needs, help achieve a healthy body weight, and reduce the risk of chronic disease. **2. Customize and enjoy nutrient-dense food and beverage choices to reflect personal preferences, cultural traditions, and budgetary considerations.** The Dietary Guidelines provide a framework of several dietary patterns intended to be customized to individual needs and preferences, as well as the foodways of the diverse cultures in the United States. **3. Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits.** Nutrient-dense foods provide vitamins, minerals, and other health-promoting components and have no or little added sugars, saturated fat, and sodium. A healthy dietary pattern consists of nutrient-dense forms of foods and beverages across all food groups, in recommended amounts, and within calorie limits. **4. Limit foods and beverages higher in added sugars, saturated fat, and sodium, and limit alcoholic beverages.** At every life stage, meeting food group recommendations, even with nutrient-dense choices, fulfills most of a person's daily calorie needs and sodium limits, with little room for extra added sugars, saturated fat, or sodium, or for alcoholic beverages.	**The Dietary Guidelines' Key Recommendations for healthy eating patterns should be applied in their entirety, given the interconnected relationship that each dietary component can have with others. They are also intended as a framework to accommodate personal preferences, cultural traditions, and budgetary considerations.** **Focus on meeting food group needs with nutrient-dense foods and beverages, and stay within calorie limits to achieve a healthy weight and reduce the risk of chronic disease.** **The core elements that make up a healthy dietary pattern include:** • Vegetables of all types—dark green; red and orange; beans, peas, and lentils; starchy; and other vegetables • Fruits, especially whole fruit • Grains, at least half of which are whole grain • Dairy, including fat-free or low-fat milk, yogurt, and cheese, and/or lactose-free versions and fortified soy beverages and yogurt as alternatives • Protein foods, including lean meats, poultry, and eggs; seafood; beans, peas, and lentils; and nuts, seeds, and soy products • Oils, including vegetable oils and oils in food, such as seafood and nuts **A healthy eating pattern limits:** • Added sugars—Less than 10% of calories per day starting at age 2. Avoid foods and beverages with added sugars for those younger than age 2. • Saturated fat—Less than 10% of calories per day starting at age 2. • Sodium—Less than 2300 mg per day—and even less for children younger than age 14. • Alcoholic beverages—Adults of legal drinking age can choose not to drink or to drink in moderation by limiting intake to 2 drinks or less in a day for men and 1 drink or less in a day for women, when alcohol is consumed. Drinking less is better for health than drinking more. There are some adults who should not drink alcohol, such as women who are pregnant. **Meet the U.S. Department of Health and Human Services' *Physical Activity Guidelines for Americans*** In tandem with the recommendations above, Americans of all ages—children, adolescents, adults, and older adults—should meet the *Physical Activity Guidelines for Americans* to help promote health and reduce the risk of chronic disease. Americans should aim to achieve and maintain a healthy body weight. The relationship between diet and physical activity contributes to calorie balance and managing body weight.

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Specific conditions and diseases, such as diabetes, other metabolic disorders, allergies, and gastrointestinal disorders, may require tailored approaches to diet. In counseling most patients, the general approach should focus on whole foods, eating patterns, and appropriate calorie balance,

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Source: Adapted from the *Dietary Guidelines for Americans, 2020–2025.* Washington, DC: U.S. Department of Agriculture and U.S. Department of Health and Human Services; 2020. Available at *https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf.*

Specific conditions and diseases, such as diabetes, other metabolic disorders, allergies, and gastrointestinal disorders, may require tailored approaches to diet. In counseling most patients, the general approach should focus on whole foods, eating patterns, and appropriate calorie balance, rather than on specific micronutrients such as electrolytes or selected vitamins. It should be remembered that most patients have difficulty understanding nutritional labels on packaged foods, with the attendant demands on numeracy and health literacy.

Dietary guidelines are published by the U.S. Department of Agriculture (USDA) and U.S. Department of Health and Human Services every 5 years, and these guidelines have undergone substantial evolution over time. The current U.S. Dietary Guidelines and Key Recommendations for 2020–2025 are summarized in Table 2-1 and emphasize the importance of healthy eating patterns for every stage of life, to avoid chronic diseases including obesity, diabetes, cancer, and cardiovascular disease. The core elements include eating patterns with nutrient-dense (rather than calorie-dense) whole foods and appropriate caloric intake to achieve and maintain healthy weight. The USDA guidelines focus on the concept of a healthy plate (rather than the prior food pyramid) for ease of counseling and adoption. Fifty percent of the plate should consist of vegetables and whole fruits, with remaining portions for whole grains and lean protein foods. When using fat for cooking, it should be done by sauteing in healthier oils (e.g., canola oil), and addition of judicious amounts of healthy raw oils (e.g., olive oil, nuts) to dishes is appropriate. Recommendations also focus on limitation of foods and beverages higher in added sugars, saturated fat, and sodium, and moderation or avoidance of alcohol intake.

The USDA guidelines focus on specific healthy eating patterns that adhere to these broad recommendations and are appropriate for ~97% of the general population. They identify a “Healthy U.S.-Style Dietary Pattern” that adheres closely to the evidence-based Dietary Approaches to Stop Hypertension (DASH) eating pattern but is customizable for different cultural or personal preferences. Alternative patterns, which vary more in emphasis than in content, include a “Healthy Mediterranean-Style Dietary Pattern” and a “Healthy Vegetarian Dietary Pattern.”

AGE- AND SEX-SPECIFIC RECOMMENDATIONS

Current dietary framework recommendations are generally similar for all life stages from ages ≥12 months, but recommended levels of caloric intake (and hence amounts of foods) differ by age, sex, and physical activity level. For example, recommended caloric intake ranges from 1000 calories/d for sedentary 2-year-old children to as high as 3200 calories/d for active 16- to 18-year-old young men. Recommended caloric intakes peak in late adolescence or early adulthood for men and women and gradually decrease over ensuing decades.

As with all lifestyle counseling aimed at behavior change, dietary approaches that partner with the patient and utilize motivational interviewing strategies and shared goals and commitments tend to work best, as described below (see “Approach to the Patient”).

Physical Activity

Similar to the approach to counseling regarding healthy eating patterns, recommendations on participation in physical activity emphasize the point that any physical activity is better than none. A simple rule of thumb for patients is: “If you are doing nothing, do something; and if you are doing something, do more, every day.” The evidence base for physical activity indicates that the marginal benefits from physical activity are greatest in advancing from no activity to low levels of moderate activity. With increasing duration and intensity of activity, there is a continued curvilinear increase in health benefits, but the marginal gains for each additional minute of moderate-to-vigorous activity slowly diminish. Thus, for adults, the recommended amount of physical activity is 150 min of moderate-intensity or 75 min of vigorous-intensity aerobic activity per week, performed in episodes of at least 5 min, and preferably spread throughout the week, plus participation in muscle-strengthening activity at least 2 days per week. Additional health benefits can be realized by engaging in physical activity beyond this amount.

In counseling patients regarding physical activity, it is important to note that sedentary time (e.g., seated at work or at home in front of electronic screens) has adverse health consequences independent of the lack of physical activity during these episodes. Therefore, even modest efforts like standing at the desk and doing gentle stretching for periods during the day may be beneficial. It is also important to emphasize that participating in a variety of aerobic activities (biking, swimming, walking, jogging, rowing, elliptical training, stair-climbing, etc.) can be beneficial and may help to avoid overuse injuries and boredom with the exercise regimen. If patients choose to participate in muscle-strengthening activities for health improvement, emphasis should be placed on weights that allow more repetitions (e.g., 3 sets of 15–20 repetitions that can be performed comfortably, with a rest period in between) and on avoiding breath-holding and straining against a closed glottis.

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

increase directly with the amount of time spent exercising, this association is substantially mitigated by training effects. Thus, patients embarking on an exercise program should be encouraged to increase the duration of aerobic exercise gradually as tolerated, aiming for episodes of at least 30 min 5

standing at the desk and doing gentle stretching for periods during the day may be beneficial. It is also important to emphasize that participating in a period in between) and on avoiding breath-holding and straining against a closed glottis.

SUDDEN CARDIAC DEATH RISK

Patients may express concerns regarding the risk of sudden cardiac death during exercise. Whereas the risk of sudden death during exercise does increase directly with the amount of time spent exercising, this association is substantially mitigated by training effects. Thus, patients embarking on an exercise program should be encouraged to increase the duration of aerobic exercise gradually as tolerated, aiming for episodes of at least 30 min 5 times a week as an ideal. Once a comfortable duration is reached, incorporating interval training periods of more intensive activity interspersed during the exercise can provide greater fitness gains.

EXTREME ENDURANCE ACTIVITIES

As with other forms of exercise, extreme endurance activities such as triathlons and marathons should be undertaken only with appropriate and graded training. Such activities tend to take a greater toll on the musculoskeletal system over time than less extreme activities, and they are also associated with measurable damage to the myocardium and greater risks for other organ damage. Athletes participating in endurance activities routinely have elevations in cardiac troponin (a specific circulating marker of myocardial cell damage and death) at the end of the race, although elevations are lower in those who are well trained. Patients and clinicians should consider the patient's overall health, specific limitations, potential for injury, and ability to train in decision-making regarding participation in endurance events.

AGE-SPECIFIC RECOMMENDATIONS

The U.S. Department of Health and Human Services' *Physical Activity Guidelines for Americans*, second edition (2018) **(Table 2-2)**, recommend that preschool-aged children (aged 3–5 years) should be physically active throughout the day in a variety of activity types to enhance growth and development. Children and adolescents aged 6–17 years should participate in ≥60 min of physical activity daily, most of which should be moderate- or vigorous-intensity aerobic activity, including vigorous, muscle-strengthening, and bone-strengthening activities at least 3 days a week each. As noted above, adults aged 18–64 years are recommended to pursue at least 150 min of moderate-intensity or 75 min of vigorous-intensity aerobic activity per week (or equivalent combinations), with at least 2 days of muscle-strengthening activities. Adults aged ≥65 years should follow the adult guidelines or be as active as possible as abilities and conditions allow. For older adults, special emphasis is also placed on multicomponent physical activity that includes balance training as well as aerobic and muscle-strengthening activities.

TABLE 2-2

Recommendations from *Physical Activity Guidelines for Americans*

Age	Recommendations
3–5 years	• Preschool-aged children (ages 3 through 5 years) should be physically active throughout the day to enhance growth and development. • Adult caregivers of preschool-aged children should encourage active play that includes a variety of activity types.

6–17 years	• It is important to provide young people opportunities and encouragement to participate in physical activities that are appropriate for their age, that are enjoyable, and that offer variety.
	• Children and adolescents ages 6 through 17 years should do 60 min (1 h) or more of moderate-to-vigorous physical activity daily:
	  ◦ Aerobic: Most of the 60 min or more per day should be either moderate- or vigorous-intensity aerobic physical activity and should include vigorous-intensity physical activity on at least 3 days a week.
	  ◦ Muscle-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include muscle-strengthening physical activity on at least 3 days a week.
	  ◦ Bone-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include bone-strengthening physical activity on at least 3 days a week.

18–64 years	• Adults should move more and sit less throughout the day. Some physical activity is better than none. Adults who sit less and do any amount of moderate-to-vigorous physical activity gain some health benefits. • For substantial health benefits, adults should do at least 150 min (2 h and 30 min) to 300 min (5 hours) a week of moderate-intensity or 75 activity a week. • Adults should also do muscle-strengthening activities of moderate or greater intensity and that involve all major muscle groups on 2 or

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Age	Recommendations
3–5 years	• Preschool-aged children (ages 3 through 5 years) should be physically active throughout the day to enhance growth and development. • Adult caregivers of preschool-aged children should encourage active play that includes a variety of activity types.

6–17 years	• It is important to provide young people opportunities and encouragement to participate in physical activities that are appropriate for their age, that are enjoyable, and that offer variety.
	• Children and adolescents ages 6 through 17 years should do 60 min (1 h) or more of moderate-to-vigorous physical activity daily:
	  ○ Aerobic: Most of the 60 min or more per day should be either moderate- or vigorous-intensity aerobic physical activity and should include vigorous-intensity physical activity on at least 3 days a week.
	  ○ Muscle-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include muscle-strengthening physical activity on at least 3 days a week.
	  ○ Bone-strengthening: As part of their 60 min or more of daily physical activity, children and adolescents should include bone-strengthening physical activity on at least 3 days a week.

18–64 years	• Adults should move more and sit less throughout the day. Some physical activity is better than none. Adults who sit less and do any amount of moderate-to-vigorous physical activity gain some health benefits. • For substantial health benefits, adults should do at least 150 min (2 h and 30 min) to 300 min (5 hours) a week of moderate-intensity or 75 min (1 h and 15 min) to 150 min (2 h and 30 min) a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic activity. Preferably, aerobic activity should be spread throughout the week. • Additional health benefits are gained by engaging in physical activity beyond the equivalent of 300 min (5 h) of moderate-intensity physical activity a week. • Adults should also do muscle-strengthening activities of moderate or greater intensity and that involve all major muscle groups on 2 or more days a week, as these activities provide additional health benefits.

≥65 years	• The key guidelines for adults also apply to older adults. In addition, the following key guidelines are just for older adults:
	  ○ As part of their weekly physical activity, older adults should do multicomponent physical activity that includes balance training as well as aerobic and muscle-strengthening activities.
	  ○ Older adults should determine their level of effort for physical activity relative to their level of fitness.
	  ○ Older adults with chronic conditions should understand whether and how their conditions affect their ability to do regular physical activity safely.
	  ○ When older adults cannot do 150 min of moderate-intensity aerobic activity a week because of chronic conditions, they should be as physically active as their abilities and conditions allow.

Moderate-intensity physical activity: Aerobic activity that increases a person's heart rate and breathing to some extent. On a scale relative to a person's capacity, moderate-intensity activity is usually a 5 or 6 on a 0 to 10 scale. Brisk walking, dancing, swimming, or bicycling on a level terrain are examples. Vigorous-intensity physical activity: Aerobic activity that greatly increases a person's heart rate and breathing. On a scale relative to a person's capacity, vigorous-intensity activity is usually a 7 or 8 on a 0 to 10 scale. Jogging, singles tennis, swimming continuous laps, or bicycling uphill are examples. Muscle-strengthening activity: Physical activity, including exercise that increases skeletal muscle strength, power, endurance, and mass. It includes strength training, resistance training, and muscular strength and endurance exercises. Bone-strengthening activity: Physical activity that produces an impact or tension force on bones, which promotes bone growth and strength. Running, jumping rope, and lifting weights are examples.

Source: Adapted from U.S. Department of Health and Human Services. *Physical Activity Guidelines for Americans, 2nd edition.* Washington, DC: U.S. Department of Health and Human Services; 2018. Available at *https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf.*

Sleep Hygiene

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

such as impaired immune function, increased pain sensitivity, and impaired cognitive performance. Conversely, achieving appropriate levels of sleep is associated with more success in weight loss, better blood pressure control among patients with hypertension, and improved mental health and

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Sleep Hygiene

Sleeping between 7 and 9 h per night appears to be optimal for health in adults aged ≥18 years. Sleeping <7 h is associated with adverse outcomes, including obesity, diabetes, elevated blood pressure, cardiovascular disease, depression, and all-cause mortality, as well as physiologic disturbances such as impaired immune function, increased pain sensitivity, and impaired cognitive performance. Conversely, achieving appropriate levels of sleep is associated with more success in weight loss, better blood pressure control among patients with hypertension, and improved mental health and performance. Regular sleep more than 9 h per night is appropriate for children and adolescents or individuals recovering from sleep deprivation or illness, but for most individuals, the effects on health are uncertain.

Patients often express concerns about the quantity and quality of their sleep. With aging, both aspects of sleep tend to decline, even without overt sleep disorders. Documentation of sleep using a sleep log may assist in understanding different types of insomnia and sleep disorders. Encouraging daily activity to promote fatigue, avoidance of eating and drinking alcohol too close to bedtime, and regular daily sleep habits may help patients achieve better sleep. Regular use of sedative medications should generally be discouraged given the high potential for dependence, addiction, and altered sleep quality.

DISORDERS OF SLEEP

The prevalence of sleep-related breathing disorders, including obstructive sleep apnea (OSA), is poorly documented. A recent systematic review suggested that that the prevalence of clinically important OSA in the general adult population may be between 9% and 38%, with higher rates in men versus women, older versus younger adults, and those with higher versus lower BMI. Patients with persistent complaints of poor sleep quality or excessive daytime somnolence or with witnessed apneic spells may benefit from screening for sleep disorders, prior to consideration of a formal sleep study. A number of clinical tools have been developed to screen for sleep apnea, including the Epworth Sleepiness Scale, the STOP (*s*noring, *t*iredness, *o*bserved apnea, high blood *p*ressure) Questionnaire, and the STOP-Bang Questionnaire (STOP plus assessment of BMI, age, neck circumference, and gender), among others. The U.S. Preventive Services Task Force found that current evidence is insufficient to assess the balance of benefits and harms of screening for OSA in asymptomatic adults owing to a lack of validation data in primary care settings. Nonetheless, the high prevalence and significant health consequences of sleep apnea suggest that clinicians should be alert for its potential presence, particularly in patients who are obese with symptoms of excessive daytime somnolence or witnessed apnea episodes. Other sleep disorders, such as restless leg syndrome, may be identified with simple history.

Weight Management

Overweight and obesity are prevalent in epidemic proportions in the United States and other industrialized nations **(Chaps. 401 and 402)**. Since 1985, the prevalence of obesity in the United States has increased from ~10 to ~35%, and the prevalence of overweight is now ~40%. Overweight and obesity disproportionately affect individuals in lower socioeconomic strata and in many underserved minority populations, including black Americans, Latino Americans, and American Indians. In all race/ethnic groups, both overweight and obesity are associated with adverse health consequences, including diabetes, certain cancers, cardiovascular diseases, and degenerative joint disease. Eating disorders such as anorexia and bulimia are much less common but pose major health consequences for affected patients and should be suspected particularly in younger women with history of rapid weight shifts or underweight status.

Weight loss is one of the most difficult preventive interventions to achieve and sustain over time. However, several key factors can assist the patient and clinician, and early referral to a dietician can be very helpful. The first therapeutic goal is to aim for weight stabilization. Many of the risks of overweight and obesity are driven more strongly by continued weight gain, rather than overweight/obese status per se. Working with the patient to find initial strategies for weight maintenance can be a successful initial step with success for many patients. For those who can progress to considering weight loss, it is critical to help the patient understand that there is no standard solution. Experimentation and documentation are key. Tools to assist patients can include food and weight logs, activity logs, and smart phone apps. Some patients respond best to structured approaches such as intermittent fasting regimens or commercial dietary programs where meals are provided. Any of these approaches can be tried with or without social group supports.

The key construct for weight loss is, of course, negative calorie balance. This is achieved through a combination of reduced caloric intake and increased physical activity. Patients may already understand, from prior weight loss attempts, what combination works best for them to achieve this. Some patients find that they cannot lose weight without increasing their exercise. For many, reduction of caloric intake is most efficient. Encouraging the patient to find what works for them is most important. The same principle holds for dietary content. Well-done feeding studies indicate that weight

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

an important start. Once weight loss is achieved, increase in activity is often required for its successful maintenance.

Tobacco Cessation

Some patients find that they cannot lose weight without increasing their exercise. For many, reduction of caloric intake is most efficient. Encouraging the patient to find what works for them is most important. The same principle holds for dietary content. Well-done feeding studies indicate that weight loss is dependent far more on the reduction of caloric intake than on the relative composition of fat, protein, and carbohydrate in the diet. There may be other medical reasons to choose one approach over another, but if not, encouraging the patient to pick one approach and document the results is an important start. Once weight loss is achieved, increase in activity is often required for its successful maintenance.

Tobacco Cessation

Escaping nicotine dependence is another major, but critical, challenge to prevention and wellness efforts (**see Chap. 454**). The addictive effects of nicotine have been well documented, with effects that can last for years after successful cessation. Assessing a patient's past history of cessation attempts and current readiness for change are key first steps in forging a successful approach. Frequent follow-up and reinforcement, as well as use of nicotine replacement therapy and other cessation-promoting medications, are additional critical elements. Recidivism is the rule, and patients should expect to resume smoking and attempt again as they journey to tobacco cessation. Electronic cigarettes have some evidence for benefit in adult smoking cessation, but their potential for use by adolescents and young adults who are not smokers represents a major public health threat for a new generation of nicotine addiction, with unknown health consequences as a result of the high doses of nicotine delivered to developing organs, including the brain. Vaping of other substances, often in association with flavoring compounds, has also been associated with pulmonary and cardiovascular damage and should be actively discouraged.

VACCINATION

One of the major advances in public health that has contributed to increases in health and longevity worldwide is the development of safe and effective vaccinations against endemic and epidemic infectious diseases **(Chap. 123)**. Patients should be counseled regarding age-appropriate vaccinations for their children and for themselves. Some individuals may be reluctant to receive a vaccination; in these cases, listening to the patient's concerns is important, followed by explanation of the benefits to the individual, their family, and their community and review of the low risk for potential harms. It is true to say that no current vaccines are ever worse than the disease they prevent, although side effects may occur rarely. Thorough knowledge of the data on side effect rates and of efficacy will aid the clinician in helping the patient make a fully informed decision.

MENTAL HEALTH AND ADDICTION

Assessment for depression and cognitive impairment is important to address when patients exhibit symptoms or they or their family members express concerns. Both of these common conditions play a major role in reducing quality of life and are high on patients' lists of concerns, even if not clearly expressed. Screening tools for depression are reviewed in **Chap. 452**. Cognitive function decline with aging or comorbid illness, including depression, should be anticipated. Assessment tools such as the General Practitioner Assessment of Cognition or the Mini-Cog™ test are widely available and effective rapid assessment tools.

Alcohol and Opioids

Alcohol dependence and abuse are common and underdiagnosed (**see Chaps. 453 and 456**). Rapid screening tools have proven efficacy for identifying patients with alcohol problems. In a systematic review, the CAGE (*c*ut down, *a*nnoyed, *g*uilty, *e*ye opener) questionnaire was most effective at identifying alcohol abuse and dependence, with reasonable sensitivity and high specificity. The present opioid epidemic in the United States presents a new and substantial public health challenge given the high potential for dependency and abuse of these drugs. Rapid screening tools are available to assist clinicians in screening for opioid dependence.

ACCIDENTS AND SUICIDE

Regular assessment of patient safety through simple questions about seat belt use, domestic violence, and gun safety in the home continues to be an important part of health promotion and wellness. Longstanding recommendations for assessment of suicidal ideation among patients with depression or a history of suicide attempts also continue to be relevant.

APPROACH TO THE PATIENT

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

importance. Much of the evaluation, counseling, and management that focus on health promotion and prevention also require engagement and buy-in from the patient in order to assist with recognition of contributing behaviors and to promote adherence to therapeutic plans. Therefore, in

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

APPROACH TO THE PATIENT

In the context of a clinical visit focused on health assessment, health promotion, and prevention, the basic skills of history-taking are of paramount importance. Much of the evaluation, counseling, and management that focus on health promotion and prevention also require engagement and buy-in from the patient in order to assist with recognition of contributing behaviors and to promote adherence to therapeutic plans. Therefore, in addition to standard history-taking, additional skills such as motivational interviewing and eliciting patient commitments and contracting may prove of significant value. The availability of additional tools to assist with screening, monitoring, and chronic management, both online and through wearable devices and mobile health technologies, is rapidly expanding, with uncertain implications for the future. Major research gaps exist in our understanding of how best to employ these newer technologies to improve health outcomes. Concepts of behavioral economics are being explored to better understand the psychology of decision-making and incentives as a means to improve lifestyle choices and adherence to treatment plans **(Chap. 481)**.

The limited time available to clinicians and patients during a wellness visit or periodic health examination (not driven by specific patient issues) makes it important to prioritize assessment and counseling for factors that affect longevity, health span, and quality of life over approaches that may have low yield, such as the annual comprehensive physical examination in an asymptomatic patient. Setting clear expectations for the content of a wellness visit may be a first step, and scheduling follow-up visits for findings or to continue indicated counseling are important steps to achieving better health outcomes.

FURTHER READING

Boulware LE et al: Systematic review: The value of the periodic health evaluation. Ann Intern Med 146:289, 2007. [PubMed: 17310053]

Dietary Guidelines for Americans, 2020–2025. Washington, DC: U.S. Department of Agriculture and U.S. Department of Health and Human Services; 2020. Available at *https://www.dietaryguidelines.gov/sites/default/files/2020-12/Dietary_Guidelines_for_Americans_2020-2025.pdf.*

Irish LA et al: The role of sleep hygiene in promoting public health: A review of empirical evidence. Sleep Med Rev 22:23, 2015. [PubMed: 25454674]

Krogsboll LT et al: General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. BMJ 345:e7191, 2012. [PubMed: 23169868]

U.S. Department of Health and Human Services: *Physical Activity Guidelines for Americans,* 2nd ed. Washington, DC: U.S. Department of Health and Human Services; 2018. Available at *https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf.*

U.S. Preventive Services Task Force webpage. Available at *https://www.uspreventiveservicestaskforce.org/uspstf/.*

Downloaded 2022-1-3 5:40 P Your IP is 141.216.78.40
Chapter 2: Promoting Good Health, Donald M. Lloyd-Jones; Kathleen M. McKibbin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 3: Vaccine Opposition and Hesitancy

Julie A. Bettinger; Hana Mitchell

INTRODUCTION

Vaccines have been recognized as one of the top public health achievements of the twentieth century. Dramatic declines in the morbidity and mortality of vaccine-preventable diseases have been observed, and the contribution of vaccines to the elimination, control, and prevention of infectious disease cannot be overstated. However, opposition and hesitancy to vaccines exist and are not new. Vaccine hesitancy has existed since Edward Jenner introduced the first vaccine against smallpox in the eighteenth century. So why did the World Health Organization rank these attitudes as one of the ten greatest threats to public health in 2019? Are current opposition and hesitancy any different from what has been seen before? Many sociologists, public health experts, and health care providers (HCPs) argue yes. Recent social and cultural trends, combined with new communication formats, have converged to create a particularly potent form of hesitancy and what some have labeled a crisis of confidence. This crisis manifests as a lack of trust in specific vaccines, vaccine programs, researchers, HCPs, the health care system, pharmaceutical companies, academics, policymakers, governments, and authority in general. **(See "Focus: COVID-19 Vaccine Hesitancy," below.)**

The roots of modern vaccine hesitancy and opposition—defined as delay or rejection of vaccines in spite of availability—vary depending on the place and the population. For some individuals and communities, pseudoscience and false claims about the safety of existing vaccines (e.g., an unsupported link between measles vaccine and autism) have driven fears, increased hesitancy, and decreased acceptance. For others, real safety events, such as the association of narcolepsy with a specific pandemic influenza vaccine (Pandemrix), have justified concerns. In a few locations (e.g., Ukraine, Pakistan), vaccine hesitancy is the result of failed health systems or even state failures. Finally, for some groups, including some fundamentalist religious groups and alternative-culture communities, vaccine hesitancy and opposition reflect exclusion from and rejection of mainstream society and allopathic health care and manifest as a deep distrust of these institutions and their HCPs. Although the genesis of modern vaccine hesitancy is multifactorial, its outcomes are uniform: a decrease in vaccine demand and uptake, a decrease in coverage by childhood and adult vaccines, and an increase in vaccine-preventable diseases, outbreaks, and epidemics of disease. Addressing this crisis and moving people from vaccine hesitancy and refusal to acceptance and active demand require intervention at multiple levels: the individual, the health system (including public health), and the state.

This chapter will define vaccine hesitancy and briefly describe its determinants and effects in North America (the United States and Canada). Physicians and other HCPs are well positioned to address the crisis of confidence many patients feel toward HCPs and the health care system. Studies demonstrate that an unambiguous, strong recommendation by trusted HCPs is most often the reason that patients, including those who are vaccine hesitant, choose to vaccinate. Strategies for counseling vaccine-hesitant and vaccine-resistant patients will be presented and examples of strong vaccine recommendations provided. Presenting strategies to increase vaccine demand at a system and policy level is beyond the scope of this chapter. While some physicians may have roles that allow them to act at this level, all physicians can act and influence their individual patients. Strategies to create active vaccine demand at the individual level alone will not solve vaccine hesitancy, but vaccine hesitancy cannot be addressed without these efforts. **For further discussion of immunization principles and vaccine use, see Chap. 123.**

VACCINE COVERAGE AND OUTBREAKS

The epidemiologic data from measles outbreaks over the past 10 years provide an interesting illustration of the effects of vaccine opposition and hesitancy. **For further discussion of measles, see Chap. 205.**

North America

Herd immunity occurs when enough individuals in a population become immune to an infectious disease, usually through vaccination, that transmission of the infection stops. The level of immunity (or level of vaccine coverage) required to confer herd immunity varies with the specific infectious disease. Because measles is a highly contagious virus, a coverage rate of 93–95% must be achieved for vaccination to confer herd immunity

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 3-1

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

transmission of the infection stops. The level of immunity (or level of vaccine coverage) required to confer herd immunity varies with the specific infectious disease. Because measles is a highly contagious virus, a coverage rate of 93–95% must be achieved for vaccination to confer herd immunity and interrupt measles transmission. National coverage estimates place one-dose measles vaccine coverage rates in 2-year-old children at 92% in the United States and 88% in Canada. In spite of these relatively high levels of coverage in young children, numerous measles outbreaks have occurred in both countries since 2010 **(Table 3-1)**.

TABLE 3-1

Measles Outbreaks in North America

Year/Place	No. of Cases	Reason
2010/Canada	70	An infected traveler to the 2010 Winter Olympics transmitted infection to an under- and unvaccinated local population in British Columbia.
2011/Canada	776	Disease was imported from France by an unvaccinated returned traveler to Quebec. The outbreak spread in a nonvaccinating religious community and outside that community. A majority of cases occurred in under- and unvaccinated persons.
2011/United States	118	Of 118 cases, 46 were in returned travelers from Europe and Asia/Pacific regions; 105 cases (89%) occurred in unvaccinated persons.
2013/United States	58	Disease was imported by a returned unvaccinated traveler from Europe. The outbreak spread in a nonvaccinating religious community in New York.
2014/Canada	433	Disease was imported from the Netherlands. The outbreak spread in a nonvaccinating religious community in British Columbia.
2014/United States	383	The outbreak occurred in nonvaccinating religious communities in Ohio.
2015/United States	147	A multistate/multicountry outbreak was linked to Disneyland amusement park. More than 80% of cases occurred in unvaccinated persons.
2015/Canada	159	Disease was imported from the United States (part of the Disneyland outbreak) by an unvaccinated traveler. The outbreak spread in a nonvaccinating religious community in Quebec.
2017/United States	75	The outbreak occurred in an under-vaccinated community in Minnesota; 95% of patients were unvaccinated.
2018/United States	375	Disease was imported by returned unvaccinated travelers from Israel. The outbreak spread in nonvaccinating religious communities in New York and New Jersey.
2019/Canada	31	Disease was imported from Vietnam by a returned traveler to British Columbia. The outbreak spread throughout local area schools in under- and unvaccinated persons and resulted in a province-wide measles mass immunization campaign for schoolchildren.
2019/United States	1282	Outbreaks occurred in 10 states; 73% of cases (~935) were linked to outbreaks in nonvaccinating religious communities in New York.

Source: Centers for Disease Control and Prevention and Public Health Agency of Canada.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Many of the outbreaks listed in Table 3-1 were ignited by unvaccinated returned travelers from areas with existing outbreaks or epidemics, who spread disease into an unvaccinated or under-vaccinated community. Many of the outbreaks were contained within the nonvaccinating community, but

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 3-1

Measles Outbreaks in North America

Year/Place	No. of Cases	Reason
2010/Canada	70	An infected traveler to the 2010 Winter Olympics transmitted infection to an under- and unvaccinated local population in British Columbia.
2011/Canada	776	Disease was imported from France by an unvaccinated returned traveler to Quebec. The outbreak spread in a nonvaccinating religious community and outside that community. A majority of cases occurred in under- and unvaccinated persons.
2011/United States	118	Of 118 cases, 46 were in returned travelers from Europe and Asia/Pacific regions; 105 cases (89%) occurred in unvaccinated persons.
2013/United States	58	Disease was imported by a returned unvaccinated traveler from Europe. The outbreak spread in a nonvaccinating religious community in New York.
2014/Canada	433	Disease was imported from the Netherlands. The outbreak spread in a nonvaccinating religious community in British Columbia.
2014/United States	383	The outbreak occurred in nonvaccinating religious communities in Ohio.
2015/United States	147	A multistate/multicountry outbreak was linked to Disneyland amusement park. More than 80% of cases occurred in unvaccinated persons.
2015/Canada	159	Disease was imported from the United States (part of the Disneyland outbreak) by an unvaccinated traveler. The outbreak spread in a nonvaccinating religious community in Quebec.
2017/United States	75	The outbreak occurred in an under-vaccinated community in Minnesota; 95% of patients were unvaccinated.
2018/United States	375	Disease was imported by returned unvaccinated travelers from Israel. The outbreak spread in nonvaccinating religious communities in New York and New Jersey.
2019/Canada	31	Disease was imported from Vietnam by a returned traveler to British Columbia. The outbreak spread throughout local area schools in under- and unvaccinated persons and resulted in a province-wide measles mass immunization campaign for schoolchildren.
2019/United States	1282	Outbreaks occurred in 10 states; 73% of cases (~935) were linked to outbreaks in nonvaccinating religious communities in New York.

Source: Centers for Disease Control and Prevention and Public Health Agency of Canada.

The vast majority (>80%) of measles cases described in Table 3-1 occurred in under- or completely unvaccinated individuals. Of note, many of these outbreaks highlight pockets of significantly under- or unvaccinated individuals that are not apparent in national vaccine coverage statistics. Moreover, many of the outbreaks listed in Table 3-1 were ignited by unvaccinated returned travelers from areas with existing outbreaks or epidemics, who spread disease into an unvaccinated or under-vaccinated community. Many of the outbreaks were contained within the nonvaccinating community, but

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

enough clusters to sustain the spread of measles. Of further concern is the number of individuals included in outbreak statistics who have had one or sometimes even two doses of vaccine and who were thought to be protected but who still end up with the disease. The assumption is that one or two

disease into an unvaccinated or under-vaccinated community. Many of the outbreaks were contained within the nonvaccinating community, but several spread to other under-vaccinated communities geographically contiguous with the outbreak community. More concerning still are the cases and outbreaks originating in communities that had not previously been identified as nonvaccinating. These cases likely highlight pockets of unvaccinated individuals who object for cultural rather than religious reasons. In the past, these nonvaccinating individuals did not exist in large enough clusters to sustain the spread of measles. Of further concern is the number of individuals included in outbreak statistics who have had one or sometimes even two doses of vaccine and who were thought to be protected but who still end up with the disease. The assumption is that one or two doses provide full disease immunity, but this is not always true. Often, individual level characteristics (age, immune compromise, etc.) affect the individual's response to the vaccine and their level of protection. In other instances, vaccine protection can wane over time, thus leaving fully immunized individuals susceptible to infection. In fact, when herd immunity breaks (i.e., the level of immunity in a community becomes too low to prevent transmission of disease), the occurrence of cases even in fully immunized persons is seen, as reflected in outbreak statistics. As a result of decreased vaccination rates and the resulting disruption of herd immunity, these individuals may become more identifiable as non-immune.

Outside North America

Although overall coverage rates may still be high in North America, they are lower in other parts of the world. In Samoa, for example, measles–mumps–rubella (MMR) vaccine coverage before a recent outbreak was 31%; in the Philippines, it was 67%. Twenty years ago, vaccine coverage was sufficiently high in some parts of the world, including Europe, that an unvaccinated traveler from a nonvaccinating community to most regions would have been protected by herd immunity at their destinations. Today that is not the case: such travelers are likely to become infected in a country with active measles transmission and return home to spread the infection into their communities and possibly beyond. Thus active measles transmission, whether at home or abroad, places individuals who rely on herd immunity (e.g., immunocompromised persons and young infants) at increased risk.

FACTORS IN VACCINE HESITANCY

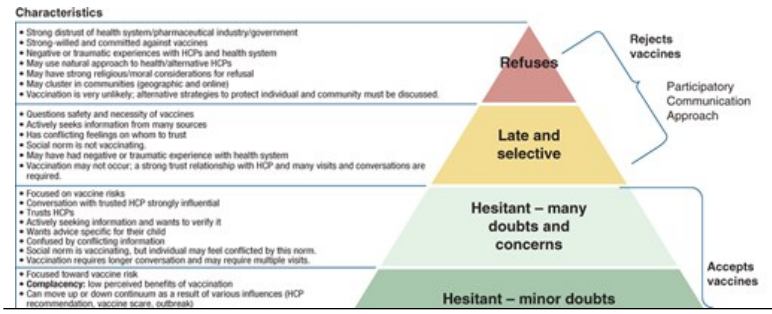
Vaccination coverage rates provide an estimate of the proportion of children or adults in the population who have been vaccinated, but they do not indicate the proportion of individuals who are vaccine hesitant. An individual may be fully vaccinated but still be hesitant about the safety and effectiveness of vaccines, or an individual may be unvaccinated as a result of access issues but may not be hesitant. Therefore, in attempts to understand a patient's lack of vaccination, it is important to distinguish persons who are hesitant and refuse vaccines from those who need assistance to access the health care system and successfully complete vaccination. To this end, an understanding of vaccine hesitancy and its determinants is needed.

Vaccine hesitancy and opposition are defined by the World Health Organization's SAGE Working Group on Vaccine Hesitancy as a "delay in acceptance or refusal of vaccines despite availability of vaccination services." The SAGE group describes vaccine hesitancy as "complex and context specific, varying across time, place, and vaccines."

It is useful to frame vaccine acceptance as a continuum pyramid, with active demand for all vaccines representing the largest group at the bottom of the pyramid and outright refusal of all vaccines depicted in the smallest group at the top. In the middle lies vaccine hesitancy, in which the degree of vaccine demand and acceptance varies. Fortunately, for disease control efforts, most individuals fall within the active-demand category or, if they are hesitant, still accept all vaccines. Hesitancy can be influenced by complacency, convenience, and confidence **(Fig. 3-1)**.

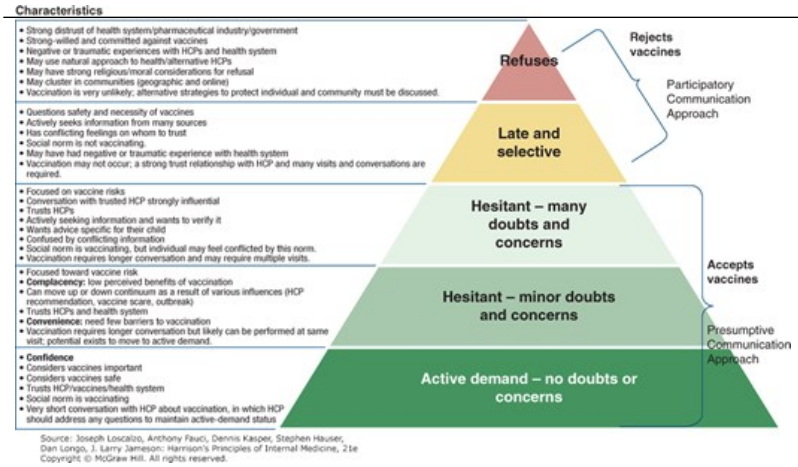
FIGURE 3-1

Vaccine acceptance continuum. HCPs, health care providers. *(Adapted from J Leask et al: BMC Pediatrics 12:154, 2012; AL Benin et al: Pediatrics 117:1532, 2006; and E Dubé, NE MacDonald: The Vaccine Book, 2016, pp. 507–528.)*



Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS



Complacency is self-satisfaction when accompanied by a lack of awareness for real dangers or deficiencies. Complacency exists in communities and individuals when the perceived risks of vaccine-preventable diseases are low and vaccination is not deemed a necessary preventive action. This attitude can apply to vaccination in general or to specific vaccines, such as influenza vaccines. Actual or perceived vaccine efficacy and effectiveness contribute to complacency. Patients who are complacent about vaccine-preventable diseases prioritize other lifestyle or health factors over vaccination. These individuals can be influenced toward vaccination by a strong recommendation from a trusted HCP or a local influenza outbreak. They can be influenced away from vaccination by a vaccine scare or misinformation on social media. Finally, the real or perceived ability of patients to take the action required for vaccination (i.e., self-efficacy) influences the role complacency plays in hesitancy and willingness to seek vaccination.

Convenience is determined by the degree to which conversations about vaccination and other services can be provided in culturally safe contexts that are convenient and comfortable for the individual. Clearly, convenience varies by community, health clinic, and even patient. Persons who are criticized or scolded for not vaccinating themselves or their children may not feel comfortable or safe accessing health services. Factors such as affordability, geographic accessibility, language, and health literacy are important considerations when evaluating the convenience of existing clinical care. Any of these factors can affect vaccine acceptance and can push a patient who has some hesitancy toward vaccinating or not vaccinating.

Confidence is based on trust in the safety and efficacy of vaccines, in the health care system that delivers vaccines (including HCPs), and in the policymakers or governments who decide which vaccines are needed and used. A continual erosion of confidence around vaccination, health systems, and governments drives today's hesitancy and has been amplified by larger social and cultural trends in medicine, parenting, and information availability.

SOCIAL AND CULTURAL TRENDS

Individualized Health Care

Over the past 30 years, the focus of medicine and health care has shifted to patient-oriented, individualized care, with an increasing emphasis on treatment and prevention options tailored to the individual patient. In vaccination programs, this shift has manifested as requests for individualized vaccine recommendations and customized immunization schedules. The increasing personalization of medicine, while positive overall, has forced public health away from a focus on the community and its common good and has created tension between individual rights and community health.

Parenting Trends

The desire for an individualized approach to medicine and vaccination reflects broader cultural trends concerning individual risk management: accordingly, the individual is to blame for bad outcomes, and public institutions cannot be trusted to manage technological (i.e., vaccine-related) risks. This viewpoint is directly linked with cultural shifts in parenting and social norms defining what it means to be a "good parent." The image of a good parent has been reframed to refer to someone whom several investigators have described as "a critical consumer of health services and products, accounting for their own individual situation as they see it with little regard for the implications of their decision on other children." The archetypical good parent no longer unquestioningly trusts HCPs and other authorities and experts. According to this social norm, "good parents" should seek individual medical advice that is tailored for their child and specific to that child's needs. While in essence not a bad thing, this norm can conflict directly with public health vaccine recommendations and schedules that are organized to maximize community health and to facilitate efficient

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Newspapers, radio, and television have been criticized for their coverage of vaccines and in particular their coverage of the alleged link between MMR

directly with public health vaccine recommendations and schedules that are organized to maximize community health and to facilitate efficient provision of care at a community level.

Traditional Media

Newspapers, radio, and television have been criticized for their coverage of vaccines and in particular their coverage of the alleged link between MMR vaccine and autism. By offering equal coverage throughout the early to mid-2000s for both the scientific evidence and unproven claims of MMR vaccine harms, traditional media outlets provided a forum and a megaphone for the spread of pseudoscience. Equal coverage leads to false equivalencies. Celebrity advocates further amplified the message via this channel. The boost that traditional media provided to active vaccine resistance and, less directly, to vaccine hesitancy has not been adequately measured but must be considered in any discussion of vaccine hesitancy. After headlines about multiple outbreaks of measles and other vaccine-preventable diseases and continued direct criticism of the equal-coverage approach, some traditional media now reject and attempt to discredit pseudoscience. The effect this stance will have on increasing vaccine confidence is unknown.

The Internet and Social Media

Approximately 90% of Americans and 91% of Canadians use the Internet, and 80% of Americans and 60% of Canadians have a social network profile. Widespread access to social media can be empowering, but it is also problematic. The Internet and social media require users to select their information sources, creating an environment described as an "echo chamber" in which individuals choose information sources harboring beliefs or opinions similar to their own and thereby reinforcing their existing views. This situation has created a new platform for further spread of vaccine *misinformation* (inaccuracies due to error) and *disinformation* (deliberate lies) and has provided a forum for vaccine-resistant individuals, including celebrities, to organize and raise funds to support their efforts. The harmful effects of Internet and social media use on vaccine hesitancy have been well documented. Vaccine hesitancy increases for parents who seek their information from the Internet. Unfortunately, public health and health care institutions have been slow to adapt to this new communication medium and to recognize its influence and impact. In this medium, personal stories and anecdotes are now viewed as data and disproportionately influence vaccine decision-making, while traditional, more authoritative, fact-based information sources are deemphasized. Centralized monitoring by jurisdiction of vaccine misinformation and disinformation, with summaries of the relevant discourses and rebuttals provided to HCPs, has been proposed as a potential way to counter the influence of social media on vaccine hesitancy. While such strategies have been applied in single jurisdictions and appear to have had some success, their applicability to a broader context is unknown. Moreover, the resources for such a coordinated response have not been made available, and individual HCPs have been left to counter popular, shifting, viral communications on their own, patient by patient.

As with traditional media, the social media landscape appears to be shifting. In 2019, the proliferation of anti-vaccination information combined with measles outbreaks in North America and increasing pressure from health leaders led large social media companies (Facebook, Instagram, Pinterest) to deemphasize anti-vaccination information by removing relevant advertisements and recommendations and decreasing their prominence in search results. While it is too soon to determine the effects of these measures, critics are skeptical that they will have the intended result of reducing vaccine misinformation and disinformation. Early evidence shows that misleading content is still widely available, with anti-vaccine advertisements now using the term "vaccine choice" to avoid censorship. More disturbingly, public health advertisements in support of vaccination have been included in social bans and removed from social media sites.

In a more grassroots effort, providers and vaccine supporters have united on social media to provide online support and evidence-based facts to providers and others who support vaccines when they are attacked digitally by anti-vaccine supporters. For example, Shots Heard Round the World (*www.shotsheard.com*) is an effort led by two U.S. pediatricians to provide advice and support for HCPs who speak out about the importance of vaccines. Such efforts harness the power of social media in ways similar to those used by vaccine opponents and may prove successful in combating vaccine hesitancy.

Given these social and cultural trends, no one should be surprised when individuals now question vaccination, express confusion about conflicting information and information sources, and feel unsure whom to trust. Their broader social context is telling them they should question everything and trust no one. This message is reinforced via misinformation and disinformation on social media. Recent vaccine-preventable disease outbreaks illustrate that effective engagement with individuals cannot be accomplished through one-way, top-down information provision (which still is often the de facto choice for health system communication), but rather requires a dialogue that takes into account the social processes surrounding individual vaccination decisions. It is at the interface between the individual and the health system in which conversations between HCPs and their patients can have the greatest impact. It is critical for all HCPs to discuss vaccines and provide strong vaccine recommendations—including HCPs who do not administer vaccines but who have established trust with their patients.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

administer vaccines but who have established trust with their patients.

APPROACH TO THE PATIENT

APPROACH TO THE PATIENT

An ideal vaccine-hesitancy intervention would result in full compliance with vaccination, the patient's satisfaction with the health care encounter, and sustained trust in the HCP's recommendations. On a programmatic level, vaccine-hesitancy interventions should be multicomponent, dialogue based, and tailored to specific under-vaccinated populations.

Communicating with vaccine-hesitant individuals can be challenging and time-consuming. HCPs may feel that vaccine-hesitant patients cast doubt on their personal and professional integrity, their authority as medical experts, and their competence as communicators. Some HCPs may be reluctant to initiate conversations about vaccination because of concerns that discussing a sensitive topic may compromise their clinical rapport with their patients. Other HCPs may believe that they have not received sufficient training to confidently recommend vaccines and answer questions. Discussing vaccines with hesitant patients, while not always easy, provides an opportunity to honor the principles of patient-centered care by demonstrating an interest in patients' opinions, engaging in dialogue, and ideally increasing patients' confidence in vaccine recommendations.

FACTORS IN EFFECTIVE VACCINE RECOMMENDATIONS

Vaccine recommendations ideally should be made within an established, trusting patient–provider relationship in which patients are comfortable asking questions and voicing concerns, even if their views on vaccines contradict the HCP's recommendations. Recommending vaccines requires both provision of information and effective communication. There is no single "best practice" for how providers should approach recommending vaccines to vaccine-hesitant individuals. In general, all vaccine recommendations should be (1) strong, making it clear that the provider supports and recommends vaccination; (2) tailored, acknowledging the vaccine attitudes and potential concerns of individual patients; (3) transparent and accurate, highlighting the benefits of vaccines while also communicating the risks; (4) supported by trustworthy information resources that patients can access and review after the clinical encounter; and (5) revisited, with repetition and reinforcement during follow-up health care encounters.

Strength of the Recommendation

HCPs should make it explicit (in the absence of medical contraindications) that vaccination based on the recommended schedule is the best option. While HCPs should take time to elicit patients' questions and address concerns, the recommendation for vaccination should be made in clear and unambiguous terms.

Tailored Communication

Vaccine hesitancy occurs on a continuum (Fig. 3-1). Therefore, it is helpful for HCPs to have some understanding of their patients' attitudes toward vaccination at the start of the health care appointment. Unfortunately, vaccine-hesitancy surveys for use as part of vaccine consultation visits have not been validated on a large scale. However, the following are some examples of questions that can be asked, depending on the setting. (1) Did you have a chance to review the vaccine leaflet we provided? Did you have any questions about it? (2) Have you ever been reluctant or hesitant about getting a vaccination for yourself or your child? If so, what were the reasons? (3) Are there other pressures in your life that prevent you from getting yourself or your child immunized on time? (4) Whom/what resources do you trust the most for information about vaccines? Whom/what resources do you trust the least?

Communication style and content for patients in the active-demand category for vaccination will be different from those for individuals who are hesitant, late and selective, or strongly inclined to refuse vaccines. Two communication styles have been proposed for vaccine recommendations. Evidence shows that a *presumptive/directive* approach ("Your child is due for MMR vaccination.") results in higher rates of vaccine uptake than a *participatory/guiding* approach ("What are your thoughts about the MMR vaccine?"). However, adopting a strictly presumptive/directive approach may alienate some patients, especially those who are higher up on the hesitancy pyramid and who may feel that they are being pressured into vaccination before their concerns have been heard and addressed. Adopting a participatory/guiding approach and clarifying receptivity to vaccines may be more suitable for hesitant individuals with many doubts and concerns, persons with a late or selective attitude, and those who are strongly inclined to refuse vaccines. In addition, a participatory/guiding approach provides an opportunity for ongoing clinical rapport and dialogue between unvaccinated or under-vaccinated patients and their HCPs, even when it does not result in immediate vaccine uptake. Regardless of which approach is used, a strong vaccine recommendation should be made at each encounter.

Transparency and Accuracy

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Childhood Vaccine Injury Act) requires HCPs to provide a copy of the current Vaccine Information Statement from the Centers for Disease Control and Prevention (CDC), which describes both benefits and risks of vaccines to an adult patient or to a child's parent/legal representative before

Transparency and Accuracy

Vaccine recommendations should be transparent, should include accurate information about both the benefits and the risks of the vaccine, and should emphasize why the benefits outweigh the risks. For example, when evidence supports an association between a vaccine and an adverse event, the occurrence of the adverse event is often very rare and the event quickly resolves **(Chap. 123).** U.S. Federal law (under the National Childhood Vaccine Injury Act) requires HCPs to provide a copy of the current Vaccine Information Statement from the Centers for Disease Control and Prevention (CDC), which describes both benefits and risks of vaccines to an adult patient or to a child's parent/legal representative before vaccination.

CDC Vaccine Information Statements should not replace a discussion with the HCP. Depending on the provider and the patient, a description of benefits and risks may include words and numbers, graphics, and personal anecdotes (e.g., why the provider vaccinates his or her own children). Personal anecdotes are powerful, and many hesitant patients seek and are influenced by them.

A discussion of benefits and risks provides an opportunity to address specific misconceptions about a particular vaccine or about vaccines overall. For example, patients may be concerned about adverse events following vaccination that are not supported by evidence, such as autism following MMR vaccination or myocardial infarction following influenza vaccination in the elderly.

Most adults—even those whose children are fully immunized—still have questions, misconceptions, or concerns about vaccines that should be addressed. A risk/benefit discussion allows HCPs to describe the vaccine safety monitoring systems in place. Providers should emphasize that vaccines are developed and approved through a highly regulated process that includes prelicensure clinical trials, review and approval by designated regulatory authorities (e.g., the U.S. Food and Drug Administration, Health Canada), strict manufacturing regulations, and ongoing postmarketing safety surveillance.

Support from Accessible Information Sources

All vaccine recommendations should be supported by additional information sources patients can assess after the health care encounter. HCPs play an important role as information intermediaries for their patients. They can navigate information (and misinformation) about vaccines and direct patients toward reliable, appropriate resources. HCPs should consider what resources will be suitable for a patient or patient population. Vaccine information resources are available in different media formats and use a combination of images and text to communicate the information to various audiences. See "Further Reading," below, for suggestions or refer to resources provided by local health authorities.

Revisiting and Reinforcement of Vaccine Recommendations

All health care encounters offer an opportunity to revisit and reinforce vaccine recommendations. Vaccine-hesitant individuals who do not accept vaccines but are willing to review information should be offered a follow-up appointment to reinforce previously made recommendations and address further questions. Vaccine-hesitant patients who accept vaccines should be seen at a follow-up appointment to confirm and document vaccine receipt (if vaccine is not given at the point of care), ascertain whether the vaccine was well tolerated, and reinforce the message about vaccine safety and effectiveness. Patients who actively demand vaccines usually do not require much follow-up other than to confirm and document the receipt of vaccine (if it is not given at the point of care) and to address additional questions or concerns arising subsequent to vaccination. Often this follow-up can be covered without an office visit.

WHAT TO SAY TO VACCINE-HESITANT PATIENTS

Engaging vaccine-hesitant individuals requires confidence, knowledge, skills, time, and creativity to tailor the approach to each individual patient. Examples for each part of the vaccine recommendation are listed in **Table 3-2.**

TABLE 3-2

Sample Vaccine Conversations

Strong vaccine recommendation
"We are headed into the flu season. Getting flu vaccine not only protects you, but it helps protect other people around you who can get very sick from flu. I strongly recommend you get your flu shot. Do you know where to get it?" "You will be turning 50 next year. This means you will be eligible for a vaccine that prevents shingles, and I strongly recommend you receive it. Have you heard about this vaccine before? Can I answer your questions about it?"
Tailored communication

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Strong vaccine recommendation
“We are headed into the flu season. Getting flu vaccine not only protects you, but it helps protect other people around you who can get very sick from flu. I strongly recommend you get your flu shot. Do you know where to get it?” “You will be turning 50 next year. This means you will be eligible for a vaccine that prevents shingles, and I strongly recommend you receive it. Have you heard about this vaccine before? Can I answer your questions about it?” “I know you are not comfortable getting vaccinated today. I do want to make it clear that I recommend vaccines because I am convinced they are the best way to protect you from some serious diseases. Is there something that would lead you to think about getting vaccinated in the future?”
Tailored communication
“I recommend that children and adults stay up to date on recommended vaccines. I see from your vaccine record that you’ve had your childhood vaccines, but you haven’t gotten any adult vaccines. I wanted to clarify whether this is because you decided not to get vaccines or something else prevented you from getting vaccinated.” “I understand that you are here for your pneumococcal vaccine. This is the best way to protect yourself and those around you from pneumonia. Do you have any questions before I give you the vaccine?” “I understand you have some concerns about vaccines. What are you most concerned about? Would you like me to explain why I recommend giving your child these vaccines?”
Transparency and accuracy
“Serious side effects can develop after MMR vaccination but are very rare. On average, 3 out of 10,000 children who get MMR vaccine will have a febrile seizure/convulsion in the days after vaccination. Febrile seizures can be frightening, but nearly all children who have a febrile seizure recover very quickly and without any long-term consequences. On the other hand, 1 out of 1000 children who get measles will develop encephalitis (brain inflammation) that not only causes seizures but can also lead to permanent damage.” “About 10 out of every 10,000 Americans who do not get vaccinated against flu die because of influenza every year, and many more are hospitalized. While flu vaccine does not prevent all cases of influenza, it is the most effective vaccine we have. By getting the vaccine, you also help protect people around you from getting sick.” “You are correct, aluminum is used in some vaccines to help the body’s immune system respond. However, aluminum is also present in food and drinking water. In fact, the amount of aluminum present in vaccines is similar to or less than what is present in breast milk or infant formulas.”
Support from accessible information sources
“Your child and other boys and girls his age will be eligible for the human papillomavirus vaccine this coming school year. Have you heard about this vaccine before? What questions do you have about it? Here’s a list of websites for parents and teenagers that explain what it is about.” “There’s a lot of information about vaccines on the internet, and a lot of that information is not based on facts. Here is a list of websites that have been reviewed by health care professionals and accurately describe benefits and risks of each vaccine. The information is written in lay language and includes helpful illustrations.”
Revisiting and reinforcement of the recommendation
“During our last visit, we talked about MMR vaccine for your son and some of the concerns you had about potential side effects. Have you had a chance to look at the take-home information I gave you? Was there anything else you would like to ask about? I recommend that we vaccinate your child today.” “During our last visit, we talked about receiving a pertussis booster during pregnancy and where you can get vaccinated. Have you had a chance to get your pertussis vaccine?” “I see that you got your vaccines at the public health clinic last week. How did it go? Did you have any questions?” “It’s possible that the symptoms you experienced after receiving the vaccine were an adverse reaction to the vaccine. I will report this to the health authority. Let’s discuss what we can do next time to prevent symptoms from occurring again.”

partnership with the CDC (vaccineinformation.org).

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Note: Specific vaccine recommendations, vaccine eligibility guidelines, and statistics used to communicate benefits and risks will vary with the health jurisdiction and the country. Several sample statements here are adapted from the Australian National Centre for Immunisation Research and Surveillance website (www.talkingaboutimmunisation.org.au). For patient vaccine information resources, see also the Immunization Action Coalition website for the public developed in partnership with the CDC (vaccineinformation.org).

OTHER CONSIDERATIONS DURING CLINICAL ENCOUNTERS

Missed Opportunities

The World Health Organization defines a missed opportunity for vaccination as "any contact with health services by an individual (child or person of any age) who is eligible for vaccination (e.g., unvaccinated or partially vaccinated and free of contraindications to vaccination), which does not result in the person receiving one or more of the vaccine doses for which he or she is eligible." HCPs who do not offer point-of-care vaccination frequently miss the opportunity to recommend vaccines to their patients. Missed opportunities for recommending and providing vaccines during routine health care encounters contribute to under-vaccination. Studies show that up to 45% of under-vaccinated children could be up to date with all age-appropriate vaccines and up to 90% of female adolescents could be up to date with human papillomavirus (HPV) vaccination if all opportunities to vaccinate were taken.

Vaccine counseling and vaccination should be incorporated into clinical care for individuals of all ages, not just young children. Because many adolescents and adults do not have regular health care follow-up, providers need to take advantage of every health care encounter to recommend and provide vaccines. For example, a visit to an emergency department, a routine follow-up visit at a diabetes clinic, or a visit planning for elective orthopedic surgery offer opportunities to inquire about the patient's vaccination status and to recommend vaccines.

HCPs should make preemptive vaccine recommendations (e.g., initiating discussions about infant vaccines during pregnancy, informing parents about HPV vaccine before their child becomes eligible). Such advance discussions may be especially helpful in identifying vaccine-hesitant patients and ensuring that they have enough time to ask questions and make decisions before vaccines are due.

HCPs should ensure that a vaccine recommendation is followed by vaccination. Providers who recommend vaccines but do not vaccinate at the point of care should inform patients where they can be vaccinated. This discussion may include information about public health clinics, travel clinics, and pharmacies or a referral to another provider. HCPs should follow up with their patients at subsequent appointments to confirm that they were vaccinated.

Adverse Events Following Vaccination

Although rare, adverse events **(Chap. 123)** may influence vaccine acceptance and willingness to be vaccinated in the future. It is important for providers to identify and follow up with all patients who experience an adverse event, regardless of the patients' vaccine attitudes prior to the event. Adverse events following vaccination should be reported to the relevant vaccine monitoring system: the U.S. Vaccine Adverse Event Reporting System or the Canadian Adverse Event Following Immunization Surveillance System.

Addressing Inequities In Vaccine Access

Discrepancies in access to health care services create inequitable access to vaccines for children and adults and contribute to under-vaccination. A U.S. study found that socially disadvantaged individuals were more likely than other persons to be under-vaccinated, in part because of a lack of access to health care services. HCPs must recognize that socially disadvantaged individuals and populations are often at greater risk of vaccine-preventable diseases (e.g., as a result of crowded living conditions, limited access to sanitation, poor nutrition, or substance abuse) and also at greater risk of being under-vaccinated because they have limited access to health care services. In addition, specific vaccines may be recommended for some socially disadvantaged populations or communities. For example, in the wake of several outbreaks of hepatitis A among the U.S. homeless population, the CDC now recommends that everyone >1 year of age experiencing homelessness receive hepatitis A vaccine.

Depending on the setting and the patient, some recommended vaccines may not be covered through public funding or private insurance coverage. HCPs should be aware of alternative funding models, such as the Vaccines for Children Program, which provides free vaccines for U.S. children (<19 years of age) with financial barriers to vaccine access. When vaccines are not publicly funded or covered by private insurance and patients perceive that

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Further Communication With Patients Who Refuse Vaccines

years of age) with financial barriers to vaccine access. When vaccines are not publicly funded or covered by private insurance and patients perceive that they cannot afford a vaccine, HCPs should not withhold a vaccine recommendation. The risks and benefits of vaccination still need to be communicated, with a strong recommendation, and the patient should be provided the opportunity to decide whether they can afford the vaccine.

Further Communication With Patients Who Refuse Vaccines

Fortunately, the proportion of people who completely refuse all vaccines and are not willing to talk to their HCP is small. Nevertheless, in some cases, attempts to initiate discussion and address vaccine refusal may be futile. When possible, HCPs should focus on the common goals of care and preserve the therapeutic relationship. Vaccine refusal should be well documented in the patient's chart. The HCP should continue with tailored communication and be open to future discussions. Vaccine demand and vaccine refusal are rarely static over time. **(See "Focus: COVID-19 Vaccine Hesitancy," below.)**

CONCLUSION

In summary, vaccine hesitancy is complex and context specific. It varies with time, place, patient, and vaccine. HCPs are well positioned to address vaccine hesitancy and should develop the skills, knowledge, and confidence to make strong vaccine recommendations to their patients.

FOCUS: COVID-19 VACCINE HESITANCY

As COVID-19 vaccines are used to control SARS-CoV-2, some individuals will have concerns about these vaccines and a proportion of the population will reject them. While worrisome, hesitancy about COVID-19 vaccines is not unexpected; it mirrors public concerns expressed about past pandemic influenza vaccines and other newly introduced vaccines. It has been established that the newness of any vaccine, be it a pandemic influenza vaccine or a COVID-19 vaccine, raises concern in a large percentage of the population. Politicization of COVID-19 vaccines raises additional issues for some patients.

Past Experience with New Vaccines

Past experience with new vaccines, including the H1N1 pandemic influenza vaccine in 2009 and the human papillomavirus vaccine in the early 2000s, provides a guide to topics that need to be addressed with regard to COVID-19 vaccines. While resistance is often framed as uncertainty about a vaccine's "newness," further discussion translates this uncertainty into concern about the new vaccine's safety. This concern encompasses both short- and long-term side effects. Frequent, acute adverse effects can be captured in clinical trial data, whereas worries about rare and long-term side effects can be addressed only by direct evidence after the initiation of a new vaccination program. In addition to queries about the overall safety of the vaccine, HCPs can expect specific questions regarding the safety of individual ingredients included in the vaccine, whether or not these ingredients are new and whether or not relevant safety data are available. Information on the incidence of common or expected health events in an unvaccinated population (i.e., background rates) over a 4-week period is helpful in distinguishing what is normal and expected from a point of concern. Studies that have examined this issue with regard to other vaccines can be used as a basis for presenting background rates of expected events in the context of COVID-19 vaccines for some groups; however, it is important to ensure that more specific background-rate information is available to HCPs with regard to the individual groups being vaccinated. HCPs, public health programs, and vaccine manufacturers can anticipate these questions and should develop answers and information to respond to them.

Specific Concerns about COVID-19 Vaccines

While some concerns can be anticipated on the basis of past experience with new vaccines, several characteristics of COVID-19 vaccines require new approaches to adequately address individual concerns, and HCPs need to educate themselves in several specific areas. First, an overwhelming amount of attention has been paid to the speed of development of COVID-19 vaccines, with some jurisdictions even skipping the usual clinical-trial steps in an effort to provide vaccine more rapidly to their populations. This situation directly increases concerns about the "newness" of the vaccine and its safety and, unfortunately, raises questions about the entire vaccine development process. Education is required to explain how a process that normally requires 5–10 years was condensed to this degree. (See Lurie et al [2020] for an excellent explanation of the COVID-19 vaccine development process.) In addition, transparency with regard to clinical trial data is required to enable scientists, HCPs, and consumers to read and understand the development and evaluation processes. The usually shrouded, proprietary development process is unsuitable if the final vaccine product is to garner public trust. Education on existing vaccine-safety monitoring systems also needs to be provided. HCPs must familiarize themselves with the vaccine development process and safety monitoring systems if they are to present this information to their patients.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

the potential for long-term adverse effects. Again, HCPs need to familiarize themselves with the new technology and develop effective messaging for their patients. Public health officials have developed resources to address this issue (see *www.cdc.gov/vaccines/covid-19/vaccinate-with-*

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

Education on existing vaccine-safety monitoring systems also needs to be provided. HCPs must familiarize themselves with the vaccine development process and safety monitoring systems if they are to present this information to their patients.

Second, several newer vaccine platforms that are being used for COVID-19 vaccines (e.g., nucleic acid–based vaccines, viral vector) have not been used in the past. This novelty exacerbates public concern about the unfamiliarity of new vaccines and further heightens misgivings about vaccine safety and the potential for long-term adverse effects. Again, HCPs need to familiarize themselves with the new technology and develop effective messaging for their patients. Public health officials have developed resources to address this issue (see *www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html*), but, even in the absence of such resources, HCPs can anticipate questions about the new technology involved and become comfortable explaining it.

Third, clinical trial safety and efficacy data were lacking for all groups initially prioritized to receive the vaccine. For example, long-term-care residents were prioritized for vaccine receipt, but clinical trial data were not available for the range of chronic health conditions that exist in older adults. While observational studies have filled some of these gaps, HCPs need to extrapolate on the basis of available evidence in considering individual patients and must make a recommendation without knowing all the answers.

Fourth, some minority and marginalized communities who have been disproportionately affected by COVID-19 express hesitancy or reject COVID-19 vaccines. For some Black, Indigenous, Latinx, and other communities, COVID-19 hesitancy stems directly from systematic discrimination, racism, and mistreatment in the health care system. Black and Indigenous communities also share a horrific legacy of unethical medical experimentation,[1] which, when combined with current discrimination and overt racism, creates a powerful climate of mistrust in HCPs, the medical system, and science.

Social and Cultural Trends

The social and cultural trends already discussed in this chapter—in particular, traditional media, the Internet, and social media—are exerting influence and pressure that did not affect the introduction of older vaccines, even the H1N1 pandemic vaccines. The media attention given to the development of transverse myelitis in one clinical-trial participant following receipt of COVID-19 vaccine is but one example of the intense media scrutiny of the vaccine development process. Unfortunately, in the United States, efforts to control COVID-19, including vaccine development, have become highly politicized. This degree of politicization has not occurred with past vaccines, so HCPs are in uncharted territory in terms of how to address it or even to understand its potential influence on vaccine acceptance. Again, individual HCPs need to navigate complex conversations with their patients and possibly their communities. Below are some suggestions that may prove helpful in formulating these conversations.

Tips for Discussion of COVID-19 Vaccines

Address Concerns about "Newness"

HCPs need to understand and be able to explain the newer vaccine platforms (mRNA, DNA, and viral vector vaccines) and to provide examples of other, older vaccines that have been developed by similar techniques. This information makes COVID-19 vaccines more familiar.

Address Concerns about Vaccine Safety

HCPs need to understand and explain how vaccines are evaluated before being approved for use and how vaccine safety is monitored after vaccines are used in the population. It is important to be honest and state that potential rare and long-term effects are not yet known, but then to speak to what is from the animal and clinical trial data and to comment on background rates for rare events. Placing potential vaccine risks in the context of known COVID-19 disease risks is helpful for some patients.

Depending on the context, explain why specific high-risk groups may have been prioritized to receive the vaccine. Patients who have been prioritized may still need a strong recommendation from an HCP to accept the vaccine. An HCP recommendation is as important here as it is for acceptance of routine vaccines. As with other vaccines, many patients' decision to accept a COVID-19 vaccine rests upon whether their HCP recommends it.

Address implicit or overt racism and systemic discrimination in the medical system and create culturally safe health care spaces. HCPs need to be aware of the legacy of discrimination, racism, and medical experimentation and the distrust it fosters in some communities. While SARS-CoV-2 has critically highlighted fractures in our health care system for minority and marginalized communities, addressing these underlying issues goes beyond addressing vaccine hesitancy and is clearly needed for all types of medical care in these communities.

Emphasize the Importance of Keeping up to Date with Other Routine Vaccines during the COVID-19 Pandemic

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

[1]The Tuskegee Syphilis Study is the most infamous example of medical experimentation in Black communities in the United States. (See Brandt [1978] for details.) Numerous examples of medical experimentation on Indigenous peoples are available. For example, a 12-year trial of an experimental

Emphasize the Importance of Keeping up to Date with Other Routine Vaccines during the COVID-19 Pandemic

These vaccines include but are not limited to seasonal influenza vaccine and the childhood primary vaccination series.

[1]The Tuskegee Syphilis Study is the most infamous example of medical experimentation in Black communities in the United States. (See Brandt [1978] for details.) Numerous examples of medical experimentation on Indigenous peoples are available. For example, a 12-year trial of an experimental bacille Calmette-Guérin vaccine for tuberculosis was conducted on Cree and Nakoda Oyadebi infants in Saskatchewan during the 1930s. (See Lux [2016] for details.)

FURTHER READING

Vaccine Hesitancy

American Academy of Pediatrics: Vaccine hesitant parents. Available at *www.aap.org/en-us/advocacy-and-policy/aap-health-initiatives/immunizations/Pages/vaccine-hesitant-parents.aspx*. Accessed October 23, 2020.

Destefano F et al: Principal controversies in vaccine safety in the United States. Clin Infect Dis 69:726, 2019. [PubMed: 30753348]

Dudley MZ et al: The state of vaccine safety science: Systematic reviews of the evidence. Lancet Infect Dis 20:e80, 2020. [PubMed: 32278359]

Immunization Action Coalition: For healthcare professionals. Available at *www.immunize.org*. Accessed October 23, 2020.

Immunization Action Coalition: For the public: Vaccine information you need. Available at *vaccineinformation.org*. Accessed October 23, 2020.

Jamison AM et al: Vaccine-related advertising in the Facebook Ad Archive. Vaccine 38:512, 2020. [PubMed: 31732327]

Leask J et al: Communicating with parents about vaccination: A framework for health professionals. BMC Pediatr 12:154, 2012. [PubMed: 22998654]

Macdonald N et al: Vaccine hesitancy: Definition, scope and determinants. Vaccine 33:4161, 2015. [PubMed: 25896383]

World Health Organization: Vaccine hesitancy survey questions related to SAGE vaccine hesitancy. Available at *www.who.int/immunization/programmes_systems/Survey_Questions_Hesitancy.pdf*. Accessed October 23, 2020.

World Health Organization: Improving vaccination demand and addressing hesitancy. Available at *www.who.int/immunization/programmes_systems/vaccine_hesitancy/en/*. Accessed October 23, 2020.

World Health Organization: Missed opportunities for vaccination (MOV) strategy. Available at *www.who.int/immunization/programmes_systems/policies_strategies/MOV/en/*. Accessed October 23, 2020.

Covid-19 Vaccine Hesitancy

Brandt AM: Racism and research: The case of the Tuskegee Syphilis Study. Hastings Cent Rep 8:21, 1978. [PubMed: 721302]

Centers for Disease Control and Prevention: Vaccinate with confidence: Strategy to reinforce confidence in Covid-19 vaccines. Available at *www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence.html*. Accessed April 5, 2021.

Lurie N et al: Developing Covid-19 vaccines at pandemic speed. N Engl J Med 382:21, 2020.

Lux MK: Separate beds: A history of Indian hospitals in Canada, 1920s–1980s. Toronto, University of Toronto Press, 2016.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Lux MK: Separate beds: A history of Indian hospitals in Canada, 1920s–1980s. Toronto, University of Toronto Press, 2016.

Mosby I et al: Medical experimentation and the roots of COVID-19 vaccine hesitancy among Indigenous Peoples in Canada. CMAJ 193:E381, 2021. [PubMed: 33627413]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 3: Vaccine Opposition and Hesitancy, Julie A. Bettinger; Hana Mitchell
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 4: Decision-Making in Clinical Medicine

Daniel B. Mark; John B. Wong

INTRODUCTION

Practicing medicine at its core requires making decisions. What makes medical practice so difficult is not only the specialized technical knowledge required but also the intrinsic uncertainty that surrounds each decision. Mastering the technical aspects of medicine alone, unfortunately, does not ensure a mastery of the practice of medicine. Sir William Osler's familiar quote "Medicine is a science of uncertainty and an art of probability" captures well this complex duality. Although the science of medicine is often taught as if the mechanisms of the human body operate with Newtonian predictability, every aspect of medical practice is infused with an element of irreducible uncertainty that the clinician ignores at her peril. Although deeply rooted in science, more than 100 years after the practice of medicine took its modern form, it remains at its core a craft, to which individual doctors bring varying levels of skill and understanding. With the exponential growth in medical literature and other technical information and an ever-increasing number of testing and treatment options, twenty-first century physicians who seek excellence in their craft must master a more diverse and complex set of skills than any of the generations that preceded them. This chapter provides an introduction to three of the pillars upon which the craft of modern medicine rests: (1) expertise in clinical reasoning (what it is and how it can be developed); (2) rational diagnostic test use and interpretation; and (3) integration of the best available research evidence with clinical judgment in the care of individual patients (evidence-based medicine [EBM]).

BRIEF INTRODUCTION TO CLINICAL REASONING

Clinical Expertise

Defining "clinical expertise" remains surprisingly difficult. Chess has an objective ranking system based on skill and performance criteria. Athletics, similarly, have ranking systems to distinguish novices from Olympians. But in medicine, after physicians complete training and pass the boards (or get recertified), no tests or benchmarks are used to identify those who have attained the highest levels of clinical performance. At each institution, there are often a few "elite" clinicians who are known for their "special problem-solving prowess" when particularly difficult or obscure cases have baffled everyone else. Yet despite their skill, even such master clinicians typically cannot explain their exact processes and methods, thereby limiting the acquisition and dissemination of the expertise used to achieve their impressive results. Furthermore, clinical virtuosity appears not to be generalizable, e.g., an expert on hypertrophic cardiomyopathy may be no better (and possibly worse) than a first-year medical resident at diagnosing and managing a patient with neutropenia, fever, and hypotension.

Broadly construed, clinical expertise encompasses not only cognitive dimensions involving the integration of disease knowledge with verbal and visual cues and test interpretation but also potentially the complex fine-motor skills necessary for invasive procedures and tests. In addition, "the complete package" of expertise in medicine requires effective communication and care coordination with patients and members of the medical team. Research on medical expertise remains sparse overall and mostly centered on diagnostic reasoning, so in this chapter, we focus primarily on the cognitive elements of clinical reasoning.

Because clinical reasoning occurs in the heads of clinicians, objective study of the process is difficult. One research method used for this area asks clinicians to "think out loud" as they receive increments of clinical information in a manner meant to simulate a clinical encounter. Another research approach focuses on how doctors should reason diagnostically, to identify remediable "errors," rather than on how they actually do reason. Much of what is known about clinical reasoning comes from empirical studies of nonmedical problem-solving behavior. Because of the diverse perspectives contributing to this area, with important contributions from cognitive psychology, medical education, behavioral economics, sociology, informatics, and decision sciences, no single integrated model of clinical reasoning exists, and not infrequently, different terms and reasoning models describe similar phenomena.

Intuitive Versus Analytic Reasoning

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

neuristics). For example, a very simple pattern that could be useful in certain situations is "black woman plus hilar adenopathy equals sarcoid." Because no effort is involved in recalling the pattern, the clinician is often unable to say how those judgments were formulated. In contrast, *Analysis*

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Intuitive Versus Analytic Reasoning

A useful contemporary model of reasoning, the dual-process theory distinguishes two general conceptual modes of thinking as fast or slow. *Intuition* (System 1) provides rapid effortless judgments from memorized associations using pattern recognition and other simplifying "rules of thumb" (i.e., heuristics). For example, a very simple pattern that could be useful in certain situations is "black woman plus hilar adenopathy equals sarcoid." Because no effort is involved in recalling the pattern, the clinician is often unable to say how those judgments were formulated. In contrast, *Analysis* (System 2), the other form of reasoning in the dual-process model, is slow, methodical, deliberative, and effortful. A student might read about causes of hilar adenopathy and from that list (e.g., **Chap. 66**), identify diseases more common in black women or examine the patient for skin or eye findings that occur with sarcoid. These dual processes, of course, represent two exemplars taken from the cognitive continuum. They provide helpful descriptive insights but very little guidance in how to develop expertise in clinical reasoning. How these idealized systems interact in different decision problems, how experts use them differently from novices, and when their use can lead to errors in judgment remain the subject of study and considerable debate.

Pattern recognition, an important part of System 1 reasoning, is a complex cognitive process that appears largely effortless. One can recognize people's faces, the breed of a dog, an automobile model, or a piece of music from just a few notes within milliseconds without necessarily being able to articulate the specific features that prompted the recognition. Analogously, experienced clinicians often recognize familiar diagnostic patterns very quickly. The key here is having a large library of stored patterns that can be rapidly accessed. In the absence of an extensive stored repertoire of diagnostic patterns, students (as well as experienced clinicians operating outside their area of expertise and familiarity) often must use the more laborious System 2 analytic approach along with more intensive and comprehensive data collection to reach the diagnosis.

The following brief patient scenarios illustrate three distinct patterns associated with hemoptysis that experienced clinicians recognize without effort:

- A 46-year-old man presents to his internist with a chief complaint of hemoptysis. An otherwise healthy, nonsmoker, he is recovering from an apparent viral bronchitis. This presentation pattern suggests that the small amount of blood-streaked sputum is due to acute bronchitis, so that a chest x-ray provides sufficient reassurance that a more serious disorder is absent.
- In the second scenario, a 46-year-old patient who has the same chief complaint but with a 100-pack-year smoking history, a productive morning cough with blood-streaked sputum, and weight loss fits the pattern of carcinoma of the lung. Consequently, along with the chest x-ray, the clinician obtains a sputum cytology examination and refers this patient for a chest CT scan.
- In the third scenario, the clinician hears a soft diastolic rumbling murmur at the apex on cardiac auscultation in a 46-year-old patient with hemoptysis who immigrated from a developing country and orders an echocardiogram as well, because of possible pulmonary hypertension from suspected rheumatic mitral stenosis.

Pattern recognition by itself is not, however, sufficient for secure diagnosis. Without deliberative systematic reflection, undisciplined pattern recognition can result in premature closure: mistakenly jumping to the conclusion that one has the correct diagnosis before all the relevant data are in. A critical second step, therefore, even when the diagnosis seems obvious, is *diagnostic verification*: considering whether the diagnosis adequately accounts for the presenting symptoms and signs and can explain all the ancillary findings. The following case based on a real clinical encounter provides an example of premature closure. A 45-year-old man presents with a 3-week history of a "flulike" upper respiratory infection (URI) including dyspnea and a productive cough. The emergency department (ED) clinician pulled out a "URI assessment form," which defines and standardizes the information gathered. After quickly acquiring the requisite structured examination components and noting in particular the absence of fever and a clear chest examination, the physician prescribed a cough suppressant for acute bronchitis and reassured the patient that his illness was not serious. Following a sleepless night at home with significant dyspnea, the patient developed nausea and vomiting and collapsed. He was brought back to the ED in cardiac arrest and was unable to be resuscitated. His autopsy showed a posterior wall myocardial infarction (MI) and a fresh thrombus in an atherosclerotic right coronary artery. What went wrong? Presumably, the ED clinician felt that the patient was basically healthy (one can be misled by the way the patient appears on examination—a patient that does not appear "sick" may be incorrectly assumed to have an innocuous illness). So, in this case, the physician, upon hearing the overview of the patient from the triage nurse, elected to use the URI assessment protocol even before starting the history, closing consideration of the broader range of possibilities and associated tests required to confirm or refute these possibilities. In particular, by concentrating on the abbreviated and focused URI protocol, the clinician failed to elicit the full dyspnea history, which was precipitated by exertion and accompanied by chest heaviness and relieved by rest, suggesting a far more serious disorder.

Heuristics or rules of thumb are a part of the intuitive system. These cognitive shortcuts provide a quick and easy path to reaching conclusions and

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

watch for the >100 cognitive biases identified to date has had any effect on the rate of diagnostic errors. In contrast, the "fast and frugal heuristics" research program explores how and when relying on simple heuristics can produce good decisions. Although many heuristics have relevance to

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Heuristics or rules of thumb are a part of the intuitive system. These cognitive shortcuts provide a quick and easy path to reaching conclusions and making choices, but when used improperly, they can lead to errors. Two major research programs have studied heuristics in a mostly nonmedical context and have reached very different conclusions about the value of these cognitive tools. The "heuristics and biases" program focuses on how these mental shortcuts can lead to incorrect judgments. So far, however, little evidence exists that educating physicians and other decision makers to watch for the >100 cognitive biases identified to date has had any effect on the rate of diagnostic errors. In contrast, the "fast and frugal heuristics" research program explores how and when relying on simple heuristics can produce good decisions. Although many heuristics have relevance to clinical reasoning, only four will be mentioned here.

When diagnosing patients, clinicians usually develop diagnostic hypotheses based on the similarity of that patient's symptoms, signs, and other data to their mental representations (memorized patterns) of the disease possibilities. In other words, clinicians pattern match to identify the diagnoses that share the most similar findings to the patient at hand. This cognitive shortcut is called the representativeness heuristic. Consider a patient with hypertension who has headache, palpitations, and diaphoresis. Based on the representativeness heuristic, clinicians might judge pheochromocytoma to be quite likely given this classic presenting symptom triad suggesting pheochromocytoma. Doing so, however, would be incorrect given that other causes of hypertension are much more common than pheochromocytoma and this triad of symptoms can occur in patients who do not have it. Thus, clinicians using the representativeness heuristic may overestimate the likelihood of a particular disease based on the presence of representative symptoms and signs, failing to account for its low underlying prevalence (i.e., the prior, or pretest, probabilities). Conversely, atypical presentations of common diseases may lead to underestimating the likelihood of a particular disease. Thus, inexperience with a specific disease and with the breadth of its presentations may also lead to diagnostic delays or errors, e.g., diseases that affect multiple organ systems, such as sarcoid or tuberculosis, may be particularly challenging to diagnose because of the many different patterns they may manifest.

A second commonly used cognitive shortcut, the availability heuristic, involves judgments based on how easily prior similar cases or outcomes can be brought to mind. For example, a clinician may recall a case from a morbidity and mortality conference in which an elderly patient presented with painless dyspnea of acute onset and was evaluated for a pulmonary cause but was eventually found to have acute MI, with the diagnostic delay likely contributing to the development of ischemic cardiomyopathy. If the case was associated with a malpractice accusation, such examples may be even more memorable. Errors with the availability heuristic arise from several sources of recall bias. Rare catastrophic outcomes become memorable cases with a clarity and force disproportionate to their likelihood for future diagnosis—for example, a patient with a sore throat eventually found to have leukemia or a young athlete with leg pain subsequently found to have an osteosarcoma—and those publicized in the media or recently experienced are, of course, easier to recall and therefore more influential on clinical judgments.

The third commonly used cognitive shortcut, the anchoring heuristic (also called conservatism or stickiness), involves insufficiently adjusting the initial probability of disease up (or down) following a positive (or negative test) when compared with Bayes' theorem, i.e., sticking to the initial diagnosis. For example, a clinician may still judge the probability of coronary artery disease (CAD) to be high despite a negative exercise perfusion test and go on to cardiac catheterization (see "Measures of Disease Probability and Bayes' Rule," below).

The fourth heuristic states that clinicians should use the simplest explanation possible that will adequately account for the patient's symptoms and findings (Occam's razor or, alternatively, the simplicity heuristic). Although this is an attractive and often used principle, it is important to remember that no biologic basis for it exists. Errors from the simplicity heuristic include premature closure leading to the neglect of unexplained significant symptoms or findings.

For complex or unfamiliar diagnostic problems, clinicians typically resort to analytic reasoning processes (System 2) and proceed methodically using the *hypothetico-deductive model of reasoning.* Based on the patient's stated reasons for seeking medical attention, clinicians develop an initial list of diagnostic possibilities in *hypothesis generation.* During the history of the present illness, the initial hypotheses evolve in *diagnostic refinement* as emerging information is tested against the mental models of the diseases being considered with diagnoses increasing and decreasing in likelihood or even being dropped from consideration as the working hypotheses of the moment. These mental models often generate additional questions that distinguish the diagnostic possibilities from one another. The focused physical examination contributes to further distinguishing the working hypotheses. Is the spleen enlarged? How big is the liver? Is it tender? Are there any palpable masses or nodules? *Diagnostic verification* involves testing the adequacy (whether the diagnosis accounts for all symptoms and signs) and coherency (whether the signs and symptoms are consistent with the underlying pathophysiologic causal mechanism) of the working diagnosis. For example, if the enlarged and quite tender liver felt on physical examination is due to acute hepatitis (the hypothesis), then certain specific liver function tests will be markedly elevated (the prediction). Should the tests come back normal, the hypothesis may have to be discarded and others reconsidered.

Although often neglected, negative findings are as important as positive ones because they reduce the likelihood of the diagnostic hypotheses under consideration. Chest discomfort that is not provoked or worsened by exertion and not relieved by rest in an active patient lowers the likelihood that

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The acuity of a patient's illness may override considerations of prevalence and the other issues described above. "Diagnostic imperatives" recognize the significance of relatively rare but potentially catastrophic conditions if undiagnosed and untreated. For example, clinicians should consider aortic

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Although often neglected, negative findings are as important as positive ones because they reduce the likelihood of the diagnostic hypotheses under consideration. Chest discomfort that is not provoked or worsened by exertion and not relieved by rest in an active patient lowers the likelihood that chronic ischemic heart disease is the underlying cause. The absence of a resting tachycardia and thyroid gland enlargement reduces the likelihood of hyperthyroidism in a patient with paroxysmal atrial fibrillation.

The acuity of a patient's illness may override considerations of prevalence and the other issues described above. "Diagnostic imperatives" recognize the significance of relatively rare but potentially catastrophic conditions if undiagnosed and untreated. For example, clinicians should consider aortic dissection routinely as a possible cause of acute severe chest discomfort. Although the typical presenting symptoms of dissection differ from those of MI, dissection may mimic MI, and because it is far less prevalent and potentially fatal if mistreated, diagnosing dissection remains a challenging diagnostic imperative **(Chap. 280)**. Clinicians taking care of acute, severe chest pain patients should explicitly and routinely inquire about symptoms suggestive of dissection, measure blood pressures in both arms for discrepancies, and examine for pulse deficits. When these are all negative, clinicians may feel sufficiently reassured to discard the aortic dissection hypothesis. If, however, the chest x-ray shows a possible widened mediastinum, the hypothesis should be reinstated and an appropriate imaging test ordered (e.g., thoracic computed tomography [CT] scan or transesophageal echocardiogram). In nonacute situations, the prevalence of potential alternative diagnoses should play a much more prominent role in diagnostic hypothesis generation.

Cognitive scientists studying the thought processes of expert clinicians have observed that clinicians group data into packets, or "chunks," that are stored in short-term or "working memory" and manipulated to generate diagnostic hypotheses. Because short-term memory is limited (classically humans can accurately repeat a list of 7 ± 2 numbers read to them), the number of diagnoses that can be actively considered in hypothesis-generating activities is similarly limited. For this reason, the cognitive shortcuts discussed above play a key role in the generation of diagnostic hypotheses, many of which are discarded as rapidly as they are formed, thereby demonstrating that the distinction between analytic and intuitive reasoning is an arbitrary and simplistic, but nonetheless useful, representation of cognition.

Research into the hypothetico-deductive model of reasoning has had difficulty identifying the elements of the reasoning process that distinguish experts from novices. This has led to a shift from examining the problem-solving process of experts to analyzing the organization of their knowledge for pattern matching as exemplars, prototypes, and illness scripts. For example, diagnosis may be based on the resemblance of a new case to patients seen previously (exemplars). As abstract mental models of disease, prototypes incorporate the likelihood of various disease features. Illness scripts include risk factors, pathophysiology, and symptoms and signs. Experts have a much larger store of exemplar and prototype cases, an example of which is the visual long-term memory of experienced radiologists. However, clinicians do not simply rely on literal recall of specific cases but have constructed elaborate conceptual networks of memorized information or models of disease to aid in arriving at their conclusions (illness scripts). That is, expertise involves an enhanced ability to connect symptoms, signs, and risk factors to one another in meaningful ways; relate those findings to possible diagnoses; and identify the additional information necessary to confirm the diagnosis.

No single theory accounts for all the key features of expertise in medical diagnosis. Experts have more knowledge about presenting symptoms of diseases and a larger repertoire of cognitive tools to employ in problem solving than nonexperts. One definition of expertise highlights the ability to make powerful distinctions. In this sense, expertise involves a working knowledge of the diagnostic possibilities and those features that distinguish one disease from another. Memorization alone is insufficient, e.g., photographic memory of a medical textbook would not make one an expert. But having access to detailed case-specific relevant information is critically important. In the past, clinicians primarily acquired clinical knowledge through their patient experiences, but now clinicians have access to a plethora of information sources. Clinicians of the future will be able to leverage the experiences of large numbers of other clinicians using electronic tools, but, as with the memorized textbook, the data alone will be insufficient for becoming an expert. Nonetheless, availability of these data removes one barrier for acquiring experience with connecting symptoms, signs, and risk factors to the possible diagnoses and identifying the additional distinguishing information necessary to confirm the diagnosis, thereby potentially facilitating the development of the working knowledge necessary for becoming an expert.

Despite all of the research seeking to understand expertise in medicine and other disciplines, it remains uncertain whether any didactic program can actually accelerate the progression from novice to expert or from experienced clinician to master clinician. Deliberate effortful practice (over an extended period of time, sometimes said to be 10 years or 10,000 practice hours) and personal coaching are two strategies often used outside medicine (e.g., music, athletics, chess) to cultivate expertise. Their use in developing medical expertise and maintaining or enhancing it has not yet been adequately explored. Some studies in medicine suggest that the most beneficial approach to education exposes students to both the signs and symptoms of specific diseases (disease pattern recognition) and, in addition, the lists of diseases that can present with specific symptoms and signs (differential diagnosis). Active learning opportunities useful for those in training include developing a personal learning system, e.g., systematically reflecting on diagnostic processes used (metacognition) and following-up to identify diagnoses and treatments for patients in their care.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The modern ideal of medical therapeutic decision-making is to "personalize" treatment recommendations. In the abstract, personalizing treatment involves combining the best available evidence about what works with an individual patient's unique features (e.g., risk factors, genomics, and

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

reflecting on diagnostic processes used (metacognition) and following-up to identify diagnoses and treatments for patients in their care.

DIAGNOSTIC VERSUS THERAPEUTIC DECISION-MAKING

The modern ideal of medical therapeutic decision-making is to "personalize" treatment recommendations. In the abstract, personalizing treatment involves combining the best available evidence about what works with an individual patient's unique features (e.g., risk factors, genomics, and comorbidities) and his or her preferences and health goals to craft an optimal treatment recommendation with the patient. Operationally, two different and complementary levels of personalization are possible: individualizing the risk of harm and benefit for the options being considered based on the specific patient characteristics (precision medicine), and personalizing the therapeutic decision process by incorporating the patient's preferences and values for the possible health outcomes. This latter process is sometimes referred to as shared decision-making and typically involves clinicians sharing their knowledge about the options and the associated consequences and trade-offs and patients sharing their health goals (e.g., avoiding a short-term risk of dying from coronary artery bypass grafting to see their grandchild get married in a few months).

Individualizing the evidence about therapy **does not** mean relying on physician impressions of benefit and harm from their personal experience. Because of small sample sizes and rare events, the chance of drawing erroneous causal inferences from one's own clinical experience is very high. For most chronic diseases, therapeutic effectiveness is only demonstrable statistically in large patient populations. It would be incorrect to infer with any certainty, for example, that treating a hypertensive patient with angiotensin-converting enzyme (ACE) inhibitors necessarily prevented a stroke from occurring during treatment, or that an untreated patient would definitely have avoided their stroke had they been treated. For many chronic diseases, a majority of patients will remain event free regardless of treatment choices; some will have events regardless of which treatment is selected; and those who avoided having an event through treatment cannot be individually identified. Blood pressure lowering, a readily observable surrogate endpoint, does not have a tightly coupled relationship with strokes prevented. Consequently, in most situations, demonstrating therapeutic effectiveness cannot rely simply on observing the outcome of an individual patient but should instead be based on large groups of patients carefully studied and properly analyzed.

Therapeutic decision-making, therefore, should be based on the best available evidence from clinical trials and well-done outcome studies. Trustworthy clinical practice guidelines that synthesize such evidence offer normative guidance for many testing and treatment decisions. However, all guidelines recognize that "one size fits all" recommendations may not apply to individual patients. Increased research into the heterogeneity of treatment effects seeks to understand how best to adjust group-level clinical evidence of treatment harms and benefits to account for the absolute level of risks faced by subgroups and even by individual patients, using, for example, validated clinical risk scores.

NONCLINICAL INFLUENCES ON CLINICAL DECISION-MAKING

More than three decades of research on variations in clinician practice patterns has identified important nonclinical forces that shape clinical decisions. These factors can be grouped conceptually into three overlapping categories: (1) factors related to an individual physician's practice, (2) factors related to practice setting, and (3) factors related to payment systems.

Factors Related to Practice Style

To ensure that necessary care is provided at a high level of quality, physicians fulfill a key role in medical care by serving as the patient's advocate. Factors that influence performance in this role include the physician's knowledge, training, and experience. Clearly, physicians cannot practice EBM if they are unfamiliar with the evidence. As would be expected, specialists generally know the evidence in their field better than do generalists. Beyond published evidence and practice guidelines, a major set of influences on physician practice can be subsumed under the general concept of "practice style." The practice style serves to define norms of clinical behavior. Differing practice styles may be based on training, personal experience, and medical evidence. Beliefs about effectiveness of different therapies and preferred patterns of diagnostic test use are examples of different facets of a practice style. For example, cardiologists evaluating patients with lower risk chest pain symptoms often conceptualize their primary diagnostic objective as maximizing the detection of ischemia. For this reason, they may strongly favor stress imaging. Internists caring for the same patients may be more comfortable with initial use of exercise ECG testing without imaging. This latter practice style focuses less on ischemia detection and more on following guideline recommendations that indicate no outcome advantage for stress imaging in this context. Cardiologist may also favor a more liberal use of coronary angiography and revascularization in patients with stable ischemic symptoms relative to general internists.

Beyond the patient's welfare, physician perceptions about the risk of a malpractice suit resulting from either an erroneous decision or a bad outcome may drive clinical decisions and create a practice referred to as defensive medicine. This practice involves ordering tests and therapies with very small

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Practice Setting Factors

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

may drive clinical decisions and create a practice referred to as defensive medicine. This practice involves ordering tests and therapies with very small marginal benefits, ostensibly to preclude future criticism should an adverse outcome occur. With conscious or unconscious awareness of a connection to the risk of litigation or to payment, however, over time, such patterns of care may become accepted as part of the practice norm, thereby perpetuating their overuse, e.g., annual cardiac exercise testing in asymptomatic patients.

Practice Setting Factors

Factors in this category relate to work systems including tasks and workflow (interruptions, inefficiencies, workload), technology (poor design or implementation, errors in use, failure, misuse), organizational characteristics (e.g., culture, leadership, staffing, scheduling), and the physical environment (e.g., noise, lighting, layout). *Physician-induced demand* is a term that refers to the repeated observation that once medical facilities and technologies become available to physicians, they will find ways to use them. Other environmental factors that can influence decision-making include the local availability of specialists for consultations and procedures; “high-tech” advanced imaging or procedure facilities such as MRI machines and proton beam therapy centers; and fragmentation of care.

Payment Systems

Economic incentives are closely related to the other two categories of practice-modifying factors. Financial issues can exert both stimulatory and inhibitory influences on clinical practice. Historically, physicians are paid on a fee-for-service, capitation, or salary basis. In fee-for-service, physicians who do more get paid more, thereby encouraging overuse, consciously or unconsciously. When fees are reduced (discounted reimbursement), clinicians tend to increase the number of services provided to maintain revenue. Capitation, in contrast, provides a fixed payment per patient per year to encourage physicians to consider a global population budget in managing individual patients and ideally reducing the use of interventions with small marginal benefit. To discourage volume-based excessive utilization, fixed salary compensation plans pay physicians the same regardless of the clinical effort expended but may provide an (unintended) incentive to see fewer patients. In recognition of the nonsustainability of continued growth in medical expenditures and the opportunity costs associated with that (funds that might be more beneficially applied to education, energy, social welfare, or defense), current efforts seek to transition to a value-based payment system to reduce overuse and to reflect benefit. Work to define how to actually tie payment to value has mostly focused so far on “pay for performance” models. High-quality clinical trial evidence for the effectiveness of these models is still mostly lacking.

INTERPRETATION OF DIAGNOSTIC TESTS

Despite impressive technological advances in medicine over the past century, uncertainty still abounds and challenges all aspects of medical decision-making. Compounding this challenge, massive information overload characterizes modern medicine. Clinicians on average subscribe to seven journals, presenting them with >2500 new articles each year, and need access to 2 million pieces of information to practice medicine. Of course, to be useful, this information must be sifted for quality and examined for applicability for integration into patient-specific care. Although computers appear to offer an obvious solution both for information management and for quantification of medical care uncertainties, many practical problems remain to be solved before computerized decision support can be routinely incorporated into the clinical reasoning process in a way that demonstrably improves the quality of care. For the present, understanding the nature of diagnostic test information can help clinicians become more efficient users of such data. The next section reviews select concepts related to diagnostic testing.

DIAGNOSTIC TESTING: MEASURES OF TEST ACCURACY

The purpose of performing a test on a patient is to reduce uncertainty about the patient’s diagnosis or prognosis in order to facilitate appropriate management. Although diagnostic tests commonly refer to laboratory (e.g., blood count) or imaging tests or procedures (e.g., colonoscopy or bronchoscopy), any information that changes a provider’s understanding of the patient’s problem qualifies as a diagnostic test. Thus, even the history and physical examination can be considered as diagnostic tests. In clinical medicine, it is common to reduce the results of a test to a dichotomous outcome, such as positive or negative, normal or abnormal. Although this simplification often suppresses useful information (such as the degree of abnormality), it facilitates illustrating some important principles of test interpretation that are described below.

The accuracy of any diagnostic test is assessed relative to a “gold standard,” where a positive gold standard test defines the patients who have disease and a negative test securely rules out disease **(Table 4-1)**. Characterizing the diagnostic performance of a new test requires identifying an appropriate population (ideally, patients representative of those in whom the new test would be used) and applying both the new and the gold standard tests to all subjects. Biased estimates of test performance occur when diagnostic accuracy is defined using an inappropriate population or one in which gold

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

who have a negative test is the false-negative rate, calculated as 1 – sensitivity. The specificity, or true-negative rate, reflects how well the new test correctly identifies patients without disease. It is the proportion of patients without disease (defined by the gold standard) who have a negative test.

population (ideally, patients representative of those in whom the new test would be used) and applying both the new and the gold standard tests to all subjects. Biased estimates of test performance occur when diagnostic accuracy is defined using an inappropriate population or one in which gold standard determination of disease status is incomplete. The accuracy of the new test in distinguishing disease from health is determined relative to the gold standard results and summarized in four estimates. The sensitivity or true-positive rate reflects how well the new test identifies patients with disease. It is the proportion of patients with disease (defined by the gold standard) who have a positive test. The proportion of patients with disease who have a negative test is the false-negative rate, calculated as 1 – sensitivity. The specificity, or true-negative rate, reflects how well the new test correctly identifies patients without disease. It is the proportion of patients without disease (defined by the gold standard) who have a negative test. The proportion of patients without disease who have positive test is the false-positive rate, calculated as 1 – specificity. In theory, a perfect test would be one with a sensitivity of 100% and a specificity of 100% and would completely distinguish patients with disease from those without it. A useful mnemonic to help remember the somewhat paradoxical relationship between what the test is best at technically versus what it is most useful for clinically is: a test with a very high sensitivity (Sn) when *negative* (N) helps *rule out* (out) disease (SnNout), and a test with a very high specificity (Sp) when *positive* (P) helps *rule in* (in) disease (SpPin).

TABLE 4-1

Measures of Diagnostic Test Accuracy

	DISEASE STATUS	
TEST RESULT	**PRESENT**	**ABSENT**
Positive	True positives (TP)	False positives (FP)
Negative	False negatives (FN)	True negatives (TN)
Test Characteristics in Patients with Disease		
True-positive rate (sensitivity) = TP/(TP + FN)		
False-negative rate = FN/(TP + FN) = 1 – true-positive rate		
Test Characteristics in Patients without Disease		
True-negative rate (specificity) = TN/(TN + FP)		
False-positive rate = FP/(TN + FP) = 1 – true-negative rate		

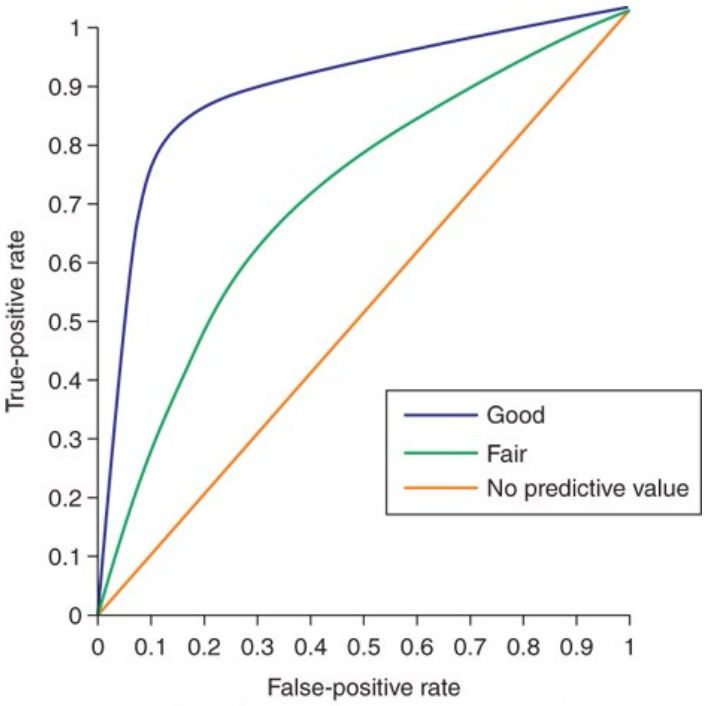
Calculating sensitivity and specificity requires selection of a threshold value or cut point above which the test is considered "positive." Making the cut point "stricter" (e.g., raising it) lowers sensitivity but improves specificity, while making it "laxer" (e.g., lowering it) raises sensitivity but lowers specificity. This dynamic trade-off between more accurate identification of subjects with disease versus those without disease is often displayed graphically as a receiver operating characteristic (ROC) curve **(Fig. 4-1)** by plotting sensitivity (*y* axis) versus 1 – specificity (*x* axis). Each point on the curve represents a potential cut point with an associated sensitivity and specificity value. The area under the ROC curve often is used as a quantitative measure of the information content of a test. Values range from 0.5 (no diagnostic information from testing at all; the test is equivalent to flipping a coin) to 1.0 (perfect test). The choice of cut point should in theory reflect the relative harms and benefits of treatment for those without versus those with disease. For example, if treatment was safe with substantial benefit, then choosing a high-sensitivity cut point (upper right of the ROC curve) for a low-risk test may be appropriate (e.g., phenylketonuria in newborns), but if treatment had substantial risk for harm, then choosing a high-specificity cut point (lower left of the ROC curve) may be appropriate (e.g., chemotherapy for cancer). The choice of cut point may also depend on the prevalence of disease, with low prevalence placing a greater emphasis on the harms of false-positive tests (e.g., HIV testing in marriage applicants) or the harms of false-negative tests (e.g., HIV testing in blood donors).

FIGURE 4-1

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

each ROC curve is a measure of the information content of the test. Thus, a larger ROC area signifies increased diagnostic accuracy.

Each receiver operating characteristic (ROC curve) illustrates a trade-off that occurs between improved test sensitivity (accurate detection of patients with disease) and improved test specificity (accurate detection of patients without disease), as the test value defining when the test turns from "negative" to "positive" is varied. A 45° line would indicate a test with no predictive value (sensitivity = specificity at every test value). The area under each ROC curve is a measure of the information content of the test. Thus, a larger ROC area signifies increased diagnostic accuracy.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

MEASURES OF DISEASE PROBABILITY AND BAYES' RULE

In the absence of perfect tests, the true disease state of the patient remains uncertain after every test. Bayes' rule provides a way to quantify the revised uncertainty using simple probability mathematics (and thereby avoid anchoring bias). It calculates the *posttest probability*, or likelihood of disease after a test result, from three parameters: the pretest probability of disease, the test sensitivity, and the test specificity. The *pretest probability* is a quantitative estimate of the likelihood of the diagnosis before the test is performed and is usually estimated from the prevalence of the disease in the underlying population (if known) or clinical context (e.g., age, sex, and type of chest pain). For some common conditions, such as CAD, existing nomograms and statistical models generate estimates of pretest probability that account for history, physical examination, and test findings. The posttest probability (also called the predictive value of the test, see below) is a recalibrated statement of the probability of the diagnosis, accounting for both pretest probability and test results. For the probability of disease following a positive test (i.e., positive predictive value), Bayes' rule is calculated as:

$$\text{Posttest probability} = \frac{\text{Pretest probability} \times \text{test sensitivity}}{\text{Pretest probability} \times \text{test sensitivity} + (1 - \text{Pretest probability}) \times (\text{false-positive test rate})}$$

For example, consider a 64-year-old woman with atypical chest pain who has a pretest probability of 0.50 and a "positive" diagnostic test result (assuming test sensitivity = 0.90 and specificity = 0.90).

$$\text{Posttest probability} = \frac{(0.50)(0.90)}{(0.50)(0.90) + (0.50)(0.10)} = 0.90$$

The term *predictive value* has often been used as a synonym for the posttest probability. Unfortunately, clinicians commonly misinterpret reported predictive values as intrinsic measures of test accuracy rather than calculated probabilities. Studies of diagnostic test performance compound the

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

in favor of the more descriptive posttest probability following a positive or a negative test result.

The nomogram version of Bayes' rule **(Fig. 4-2)** helps us to understand at a conceptual level how it estimates the posttest probability of disease. In

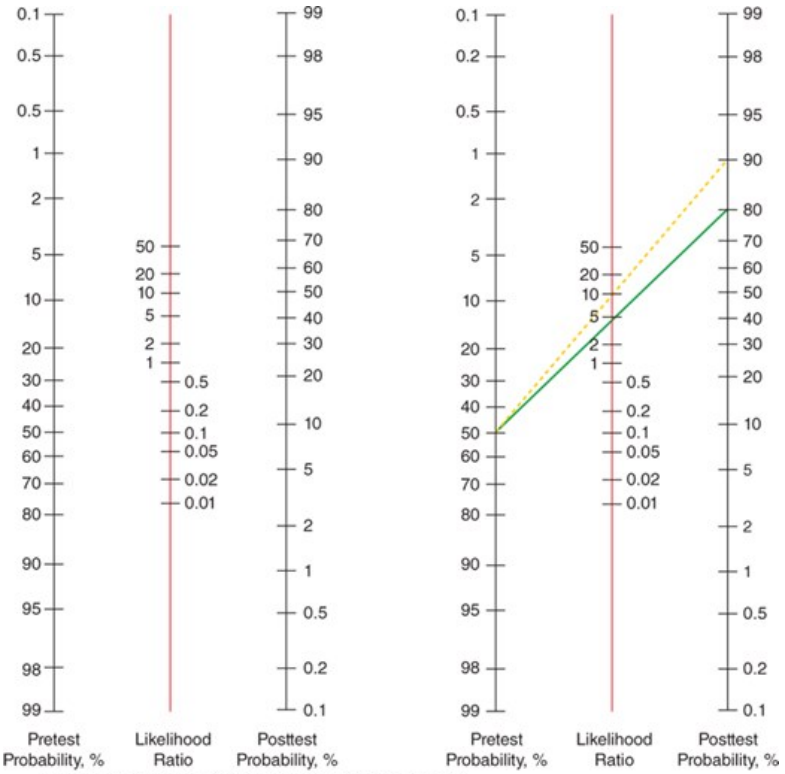
ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

The term *predictive value* has often been used as a synonym for the posttest probability. Unfortunately, clinicians commonly misinterpret reported predictive values as intrinsic measures of test accuracy rather than calculated probabilities. Studies of diagnostic test performance compound the confusion by calculating predictive values from the same sample used to measure sensitivity and specificity. Such calculations are misleading unless the test is applied subsequently to populations with exactly the same disease prevalence. For these reasons, the term *predictive value* is best avoided in favor of the more descriptive posttest probability following a positive or a negative test result.

The nomogram version of Bayes' rule **(Fig. 4-2)** helps us to understand at a conceptual level how it estimates the posttest probability of disease. In this nomogram, the impact of the diagnostic test result is summarized by the likelihood ratio, which is defined as the ratio of the probability of a given test result (e.g., "positive" or "negative") in a patient with disease to the probability of that result in a patient without disease, thereby providing a measure of how well the test distinguishes those with from those without disease.

FIGURE 4-2

Nomogram version of Bayes' theorem used to predict the posttest probability of disease (right-hand scale) using the pretest probability of disease (left-hand scale) and the likelihood ratio for a positive or a negative test (middle scale). See text for information on calculation of likelihood ratios. To use, place a straightedge connecting the pretest probability and the likelihood ratio and read off the posttest probability. The right-hand part of the figure illustrates the value of a positive exercise treadmill test (likelihood ratio 4, *green line*) and a positive exercise thallium single-photon emission CT perfusion study (likelihood ratio 9, broken *yellow line*) in a patient with a pretest probability of coronary artery disease of 50%. *(Adapted from Centre for Evidence-Based Medicine: Likelihood ratios. Available at http://www.cebm.net/likelihood-ratios/.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The *likelihood ratio for a positive test* is calculated as the ratio of the true-positive rate to the false-positive rate (or sensitivity/[1 – specificity]). For example, a test with a sensitivity of 0.90 and a specificity of 0.90 has a likelihood ratio of 0.90/(1 – 0.90), or 9. Thus, for this hypothetical test, a "positive" result is 9 times more likely in a patient with the disease than in a patient without it. Most tests in medicine have likelihood ratios for a positive result between 1.5 and 20. Higher values are associated with tests that more substantially increase the posttest likelihood of disease. A very high likelihood ratio positive (>10) usually implies high specificity, so a positive high specificity test helps "rule in" disease (the "SpPin" mnemonic introduced earlier). If sensitivity is excellent but specificity is less so, the likelihood ratio positive will be reduced substantially (e.g., with a 90% sensitivity but a 55% specificity, the likelihood ratio positive is 2.0).

The corresponding *likelihood ratio for a negative test* is the ratio of the false-negative rate to the true-negative rate (or [1 – sensitivity]/specificity).

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

that a negative result is about one-tenth as likely in patients with disease than in those without disease (or about 10 times more likely in those without disease than in those with disease).

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

The corresponding *likelihood ratio for a negative test* is the ratio of the false-negative rate to the true-negative rate (or [1 – sensitivity]/specificity). Lower likelihood ratio negative values more substantially lower the posttest likelihood of disease. A very low likelihood ratio negative (falling below 0.10) usually implies high sensitivity, so a negative high sensitivity test helps "rule out" disease (the SnNout mnemonic). The hypothetical test considered above with a sensitivity of 0.9 and a specificity of 0.9 would have a likelihood ratio for a negative test result of (1 – 0.9)/0.9, or 0.11, meaning that a negative result is about one-tenth as likely in patients with disease than in those without disease (or about 10 times more likely in those without disease than in those with disease).

APPLICATIONS TO DIAGNOSTIC TESTING IN CAD

Consider two tests commonly used in the diagnosis of CAD: an exercise treadmill and an exercise single-photon emission CT (SPECT) myocardial perfusion imaging test **(Chap. 241)**. A positive treadmill ST-segment response has an average sensitivity of ~60% and an average specificity of ~75%, yielding a likelihood ratio positive of 2.4 (0.60/[1 – 0.75]) (consistent with modest discriminatory ability because it falls between 2 and 5). For a 41-year-old man with nonanginal pain and a 10% pretest probability of CAD, the posttest probability of disease after a positive result rises to only ~30%. For a 60-year-old woman with typical angina and a pretest probability of CAD of 80%, a positive test result raises the posttest probability of disease to ~95%.

In contrast, exercise SPECT myocardial perfusion test is more accurate for diagnosis of CAD. For simplicity, assume that the finding of a reversible exercise-induced perfusion defect has both a sensitivity and a specificity of 90% (a bit higher than reported), yielding a likelihood ratio for a positive test of 9.0 (0.90/[1 – 0.90]) (consistent with intermediate discriminatory ability because it falls between 5 and 10). For the same 10% pretest probability patient, a positive test raises the probability of CAD to 50% (Fig. 4-2). However, despite the differences in posttest probabilities between these two tests (30 vs 50%), the more accurate test may not improve diagnostic likelihood enough to change patient management (e.g., decision to refer to cardiac catheterization) because the more accurate test has only moved the physician from being fairly certain that the patient did not have CAD to a 50:50 chance of disease. In a patient with a pretest probability of 80%, exercise SPECT test raises the posttest probability to 97% (compared with 95% for the exercise treadmill). Again, the more accurate test does not provide enough improvement in posttest confidence to alter management, and neither test has improved much on what was known from clinical data alone.

In general, positive results with an accurate test (e.g., likelihood ratio for a positive test of 10) when the pretest probability is low (e.g., 20%) do not move the posttest probability to a range high enough to rule in disease (e.g., 80%). In screening situations, pretest probabilities are often particularly low because patients are asymptomatic. In such cases, specificity becomes especially important. For example, in screening first-time female blood donors without risk factors for HIV, a positive test raised the likelihood of HIV to only 67% despite a specificity of 99.995% because the prevalence was 0.01%. Conversely, with a high pretest probability, a negative test may not rule out disease adequately if it is not sufficiently sensitive. Thus, the largest change in diagnostic likelihood following a test result occurs when the clinician is most uncertain (i.e., pretest probability between 30 and 70%). For example, in patients with a pretest probability for CAD of 50%, a positive exercise treadmill test moves the posttest probability to 80% and a positive exercise SPECT perfusion test moves it to 90% (Fig. 4-2).

As presented above, Bayes' rule employs a number of important simplifications that should be considered. First, few tests provide only "positive" or "negative" results. Many tests have multidimensional outcomes (e.g., extent of ST-segment depression, exercise duration, and exercise-induced symptoms with exercise testing). Although Bayes' theorem can be adapted to this more detailed test result format, it is computationally more complex to do so. Similarly, when multiple sequential tests are performed, the posttest probability may be used as the pretest probability to interpret the second test. However, this simplification assumes conditional independence—that is, that the results of the first test do not affect the likelihood of the second test result—and this is often not true.

Finally, many texts assert that sensitivity and specificity are prevalence-independent parameters of test accuracy. This statistically useful assumption, however, is often incorrect. A treadmill exercise test, for example, has a sensitivity of ~30% in a population of patients with one-vessel CAD, whereas its sensitivity in patients with severe three-vessel CAD approaches 80%. Thus, the best estimate of sensitivity to use in a particular decision may vary, depending on the severity of disease in the local population. A hospitalized, symptomatic, or referral population typically has a higher prevalence of disease and, in particular, a higher prevalence of more advanced disease than does an outpatient population. Consequently, test sensitivity will likely be higher in hospitalized patients and test specificity higher in outpatients.

STATISTICAL PREDICTION MODELS

Bayes' rule, when used as presented above, is useful in studying diagnostic testing concepts, but predictions based on multivariable statistical models can more accurately address these more complex problems by simultaneously accounting for additional relevant patient characteristics. In particular,

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

the limited data that clinicians can manage in their heads or with Bayes' rule. However, despite this strength, prediction models are usually too complex computationally to use without a calculator or computer. Guideline-driven treatment recommendations based on statistical prediction

can more accurately address these more complex problems by simultaneously accounting for additional relevant patient characteristics. In particular, these models explicitly account for multiple, even possibly overlapping, pieces of patient-specific information and assign a relative weight to each on the basis of its unique independent contribution to the prediction in question. For example, a logistic regression model to predict the probability of CAD ideally considers all the relevant independent factors from the clinical examination and diagnostic testing and their relative importance instead of the limited data that clinicians can manage in their heads or with Bayes' rule. However, despite this strength, prediction models are usually too complex computationally to use without a calculator or computer. Guideline-driven treatment recommendations based on statistical prediction models available online, e.g., the American College of Cardiology/American Heart Association risk calculator for primary prevention with statins and the CHA_2DS_2-VASC calculator for anticoagulation for atrial fibrillation, have generated more widespread usage. When electronic health records (EHRs) will provide sufficient platform support to allow for routine use of predictive models in clinical practice and increase their impact on clinical encounters and outcomes remains uncertain.

One reason for limited clinical use is that, to date, only a handful of prediction models have been validated sufficiently (for example, Wells criteria for pulmonary embolism; **Table 4-2**). The importance of independent validation in a population separate from the one used to develop the model cannot be overstated. An unvalidated prediction model should be viewed with the skepticism appropriate for any new drug or medical device that has not had rigorous clinical trial testing.

TABLE 4-2

Wells Clinical Prediction Rule for Pulmonary Embolism (PE)

CLINICAL FEATURE	POINTS
Clinical signs of deep-vein thrombosis	3
Alternative diagnosis is less likely than PE	3
Heart rate >100 beats/min	1.5
Immobilization ≥3 days or surgery in previous 4 weeks	1.5
History of deep-vein thrombosis or pulmonary embolism	1.5
Hemoptysis	1
Malignancy (with treatment within 6 months) or palliative	1
INTERPRETATION	
Score >6.0	High
Score 2.0–6.0	Intermediate
Score <2.0	Low

When statistical survival models in cancer and heart disease have been compared directly with clinicians' predictions, the survival models have been found to be more consistent, as would be expected, but not always more accurate. On the other hand, comparison of clinicians with websites and apps that generate lists of possible diagnoses to help patients with self-diagnosis found that physicians outperformed the currently available programs. For students and less-experienced clinicians, the biggest value of diagnostic decision support may be in extending diagnostic possibilities and triggering "rational override," but their impact on knowledge, information-seeking, and problem-solving needs additional research.

FORMAL DECISION SUPPORT TOOLS

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Conceptually, computers offer several levels of potentially useful support for clinicians. At the most basic level, they provide ready access to vast reservoirs of information, which may, however, be quite difficult to sort through to find what is needed. At higher levels, computers can support care

DECISION SUPPORT SYSTEMS

Over the past 50 years, many attempts have been made to develop computer systems to aid clinical decision-making and patient management. Conceptually, computers offer several levels of potentially useful support for clinicians. At the most basic level, they provide ready access to vast reservoirs of information, which may, however, be quite difficult to sort through to find what is needed. At higher levels, computers can support care management decisions by making accurate predictions of outcome, or can simulate the whole decision process, and provide algorithmic guidance. Computer-based predictions using Bayesian or statistical regression models inform a clinical decision but do not actually reach a "conclusion" or "recommendation." Machine learning methods are being applied to pattern recognition tasks such as the examination of skin lesions and the interpretation of x-rays. Artificial intelligence (AI) systems attempt to simulate or replace human reasoning with a computer-based analogue. Natural language processing allows the system to access and process large amounts of data, both from the EHR and from the medical literature. To date, such approaches have achieved only limited success. The most prominent example, IBM's Watson program, introduced publicly in 2011, has yet to produce persuasive evidence of clinical decision support utility. Reminder or protocol-directed systems do not make predictions but use existing algorithms, such as guidelines or appropriate utilization criteria, to direct clinical practice. In general, however, decision support systems have so far had little impact on practice. Reminder systems built into EHRs have shown the most promise, particularly in correcting drug dosing and promoting adherence to guidelines. Checklists may also help avoid or reduce errors.

DECISION ANALYSIS

Compared with the decision support methods discussed earlier, decision analysis represents a normative prescriptive approach to decision-making in the face of uncertainty. Its principal application is in complex decisions. For example, public health policy decisions often involve *trade-offs* in length versus quality of life, benefits versus resource use, population versus individual health, and *uncertainty* regarding efficacy, effectiveness, and adverse events as well as *values* or preferences regarding mortality and morbidity outcomes.

One recent analysis using this approach involved the optimal screening strategy for breast cancer, which has remained controversial, in part because a randomized controlled trial to determine when to begin screening and how often to repeat screening mammography is impractical. In 2016, the National Cancer Institute–sponsored Cancer Intervention and Surveillance Network (CISNET) examined eight strategies differing by whether to initiate mammography screening at age 40, 45, or 50 years and whether to screen annually, biennially, or annually for women in their forties and biennially thereafter (hybrid). The six simulation models found biennial strategies to be the most efficient for average-risk women. Biennial screening for 1000 women from age 50–74 years versus no screening avoided seven breast cancer deaths. Screening annually from age 40–74 years avoided three additional deaths but required 20,000 additional mammograms and yielded 1988 more false-positive results. Factors that influenced the results included patients with a 2–4-fold higher risk for developing breast cancer in whom annual screening from age 40–74 years yielded similar benefits as biennial screening from age 50–74. For average-risk patients with moderate or severe comorbidities, screening could be stopped earlier, at age 66–68 years.

This analysis involved six models that reproduced epidemiologic trends and a screening trial result, accounted for digital technology and treatments advances, and considered quality of life, risk factors, breast density, and comorbidity. It provided novel insights into a public health problem in the absence of a randomized clinical trial and helped weigh the pros and cons of such a health policy recommendation. Although such models have been developed for selected clinical problems, their benefit and application to individual real-time clinical management has yet to be demonstrated.

DIAGNOSIS AS AN ELEMENT OF QUALITY OF CARE

High-quality medical care begins with accurate diagnosis. The incidence of diagnostic errors has been estimated by a variety of methods including postmortem examinations, medical record reviews, and medical malpractice claims, with each yielding complementary but different estimates of this quality of care patient-safety problem. In the past, diagnostic errors tended to be viewed as a failure of individual clinicians. The modern view is that they are mostly a system of care deficiencies. Current estimates suggest that nearly everyone will experience at least one diagnostic error in their lifetime, leading to mortality, morbidity, unnecessary tests and procedures, costs, and anxiety.

Solutions to the "diagnostic errors as a system of care" problem have focused on system-level approaches, such as decision support and other tools integrated into EHRs. The use of checklists has been proposed as a means of reducing some of the cognitive errors discussed earlier in the chapter, such as premature closure. While checklists have been shown to be useful in certain medical contexts, such as operating rooms and intensive care units, their value in preventing diagnostic errors that lead to patient adverse events remains to be shown.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Clinical medicine is defined traditionally as a practice combining medical knowledge (including scientific evidence), intuition, and judgment in the care of patients (Chap. 1). Evidence-based medicine (EBM) updates this construct by placing much greater emphasis on the processes by which clinicians

such as premature closure. While checklists have been shown to be useful in certain medical contexts, such as operating rooms and intensive care units, their value in preventing diagnostic errors that lead to patient adverse events remains to be shown.

EVIDENCE-BASED MEDICINE

Clinical medicine is defined traditionally as a practice combining medical knowledge (including scientific evidence), intuition, and judgment in the care of patients (Chap. 1). Evidence-based medicine (EBM) updates this construct by placing much greater emphasis on the processes by which clinicians gain knowledge of the most up-to-date and relevant clinical research to determine for themselves whether medical interventions alter the disease course and improve the length or quality of life. The phrase "evidence-based medicine" is now used so often and in so many different contexts that many practitioners are unaware of its original meaning. The intention of the EBM program, as described in the early 1990s by its founding proponents at McMaster University, becomes clearer through an examination of its four key steps:

1. Formulating the management question to be answered
2. Searching the literature and online databases for applicable research data
3. Appraising the evidence gathered with regard to its validity and relevance
4. Integrating this appraisal with knowledge about the unique aspects of the patient (including the patient's preferences about the possible outcomes)

The process of searching the world's research literature and appraising the quality and relevance of studies can be time-consuming and requires skills and training that most clinicians do not possess. In a busy clinical practice, the work required is also logistically not feasible. This has led to a focus on finding recent systematic overviews of the problem in question as a useful shortcut in the EBM process. Systematic reviews are regarded by some as the highest level of evidence in the EBM hierarchy because they are intended to comprehensively summarize the available evidence on a particular topic. To avoid the potential biases found in narrative review articles, predefined reproducible explicit search strategies and inclusion and exclusion criteria seek to find all of the relevant scientific research and grade its quality. The prototype for this kind of resource is the Cochrane Database of Systematic Reviews. When appropriate, a meta-analysis is used to quantitatively summarize the systematic review findings (discussed further below).

Unfortunately, systematic reviews are not uniformly the acme of the EBM process they were initially envisioned to be. In select circumstances, they can provide a much clearer picture of the state of the evidence than is available from any individual clinical report, but their value is less clear when only a few trials are available, when trials and observational studies are mixed, or when the evidence base is only observational. They cannot compensate for deficiencies in the underlying research available, and many are created without the requisite clinical insights. The medical literature is now flooded with systematic reviews of varying quality and clinical utility. The peer review system has, unfortunately, not proved to be an effective arbiter of quality of these papers. Therefore, systematic reviews should be used with circumspection in conjunction with selective reading of some of the best empirical studies.

SOURCES OF EVIDENCE: CLINICAL TRIALS AND REGISTRIES

The notion of learning from observation of patients is as old as medicine itself. Over the past 50 years, physicians' understanding of how best to turn raw observation into useful evidence has evolved considerably. Medicine has received a hard refresher lesson in this process from COVID-19 pandemic. Starting in the spring of 2020, case reports, personal and institutional anecdotal experience, and small single-center case series started appearing in the peer-reviewed literature and within months turned into a flood of confusing and often contradictory evidence. Observational reports of treatments for COVID-19 fueled the confusion. Despite >40,000 publications appearing in the first 7 months of the pandemic, an enormous amount of uncertainty around prevention, diagnosis, treatment, and prognosis of the disease remained. Many of the early 2020 publications were either small observational series or reviews of published series, neither of which can resolve the key uncertainties clinicians need to address in caring for these patients. These small observational studies often have substantial limitations in validity and generalizability, and although they may generate important hypotheses or be the first reports of adverse events or therapeutic benefit, they have no role in formulating modern standards of practice. The major tools used to develop reliable evidence consist of randomized clinical trials supplemented strategically by large (high-quality) observational registries. A registry or database typically is focused on a disease or syndrome (e.g., different types of cancer, acute or chronic CAD, pacemaker capture, or chronic heart failure), a clinical procedure (e.g., bone marrow transplantation, coronary revascularization), or an administrative process (e.g., claims data used for billing and reimbursement).

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

(e.g., claims data). Advantages of observational data include the inclusion of a broader population as encountered in practice than is typically represented in clinical trials because of their restrictive inclusion and exclusion criteria. In addition, observational data provide primary evidence for

By definition, in observational data, the investigator does not control patient care. Carefully collected prospective observational data, however, can at times achieve a level of evidence quality approaching that of major clinical trial data. At the other end of the spectrum, data collected retrospectively (e.g., chart review) are limited in form and content to what previous observers recorded and may not include the specific research data being sought (e.g., claims data). Advantages of observational data include the inclusion of a broader population as encountered in practice than is typically represented in clinical trials because of their restrictive inclusion and exclusion criteria. In addition, observational data provide primary evidence for research questions when a randomized trial cannot be performed. For example, it would be difficult to randomize patients to test diagnostic or therapeutic strategies that are unproven but widely accepted in practice, and it would be unethical to randomize based on sex, racial/ethnic group, socioeconomic status, or country of residence or to randomize patients to a potentially harmful intervention, such as smoking or deliberately overeating to develop obesity.

A well-done prospective observational study of a particular management strategy differs from a well-done randomized clinical trial most importantly by its lack of protection from treatment selection bias. The use of observational data to compare diagnostic or therapeutic strategies assumes that sufficient uncertainty and heterogeneity exists in clinical practice to ensure that similar patients will be managed differently by diverse physicians. In short, the analysis assumes that a sufficient element of randomness (in the sense of disorder rather than in the formal statistical sense) exists in clinical management. In such cases, statistical models attempt to adjust for important imbalances to "level the playing field" so that a fair comparison among treatment options can be made. When management is clearly not random (e.g., all eligible left main CAD patients are referred for coronary bypass surgery), the problem may be too confounded (biased) for statistical correction, and observational data may not provide reliable evidence.

In general, the use of concurrent controls is vastly preferable to that of historical controls. For example, comparison of current surgical management of left main CAD with medically treated patients with left main CAD during the 1970s (the last time these patients were routinely treated with medicine alone) would be extremely misleading because "medical therapy" has substantially improved in the interim.

Randomized controlled clinical trials include the careful prospective design features of the best observational data studies but also include the use of random allocation of treatment. This design provides the best protection against measured and unmeasured confounding due to treatment selection bias (a major aspect of internal validity). However, the randomized trial may not have good external validity (generalizability) if the process of recruitment into the trial resulted in the exclusion of many potentially eligible subjects or if the nominal eligibility for the trial describes a very heterogeneous population.

Consumers of medical evidence need to be aware that randomized trials vary widely in their quality and applicability to practice. The process of designing such a trial often involves many compromises. For example, trials designed to gain U.S. Food and Drug Administration (FDA) approval for an investigational drug or device must fulfill regulatory requirements (such as the use of a placebo control) that may result in a trial population and design that differ substantially from what practicing clinicians would find most useful.

META-ANALYSIS

The Greek prefix *meta* signifies something at a later or higher stage of development. Meta-analysis is research that combines and summarizes the available evidence quantitatively. Although it is used to examine nonrandomized studies, meta-analysis is most useful for summarizing all available randomized trials examining a particular therapy used in a specific clinical context. Ideally, unpublished trials should be identified and included to avoid publication bias (i.e., missing "negative" trials that may not be published). Furthermore, the best meta-analyses obtain and analyze individual patient-level data from all trials rather than using only the summary data from published reports. Nonetheless, not all published meta-analyses yield reliable evidence for a particular problem, so their methodology should be scrutinized carefully to ensure proper study design and analysis. The results of a well-done meta-analysis are likely to be most persuasive if they include at least several large-scale, properly performed randomized trials. Meta-analysis can especially help detect benefits when individual trials are inadequately powered (e.g., the benefits of streptokinase thrombolytic therapy in acute MI demonstrated by ISIS-2 in 1988 were evident by the early 1970s through meta-analysis). However, in cases in which the available trials are small or poorly done, meta-analysis should not be viewed as a remedy for deficiencies in primary trial data or trial design.

Meta-analyses typically focus on summary measures of relative treatment benefit, such as odds ratios or relative risks. Clinicians should also examine what absolute risk reduction (ARR) can be expected from the therapy. A metric of absolute treatment benefit that is frequently reported is the number needed to treat (NNT) to prevent one adverse outcome event (e.g., death, stroke). NNT should not be interpreted literally as a causal statement. NNT is simply 1/ARR. For example, if a hypothetical therapy reduced mortality rates over a 5-year follow-up by 33% (the relative treatment benefit) from 12% (control arm) to 8% (treatment arm), the ARR would be 12% – 8% = 4% and the NNT would be 1/.04, or 25. This does not mean literally that 1 patient benefits and 24 do not. However, it can be conceptualized as an informal measure of treatment efficiency. If the hypothetical treatment was applied to

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

from different studies should account for the duration of follow-up used to create each estimate. In addition, the NNT concept assumes a homogeneity in response to treatment that may not be accurate. The NNT is simply another way of summarizing the absolute treatment difference and does not provide any unique information.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

(control arm) to 8% (treatment arm), the ARR would be 12% – 8% = 4% and the NNT would be 1/.04, or 25. This does not mean literally that 1 patient benefits and 24 do not. However, it can be conceptualized as an informal measure of treatment efficiency. If the hypothetical treatment was applied to a lower-risk population, say, with a 6% 5-year mortality, the 33% relative treatment benefit would reduce absolute mortality by 2% (from 6% to 4%), and the NNT for the same therapy in this lower-risk group of patients would be 50. Although not always made explicit, comparisons of NNT estimates from different studies should account for the duration of follow-up used to create each estimate. In addition, the NNT concept assumes a homogeneity in response to treatment that may not be accurate. The NNT is simply another way of summarizing the absolute treatment difference and does not provide any unique information.

CLINICAL PRACTICE GUIDELINES

Per the 1990 Institute of Medicine definition, clinical practice guidelines are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances." This definition emphasizes several crucial features of modern guideline development. First, guidelines are created by using the tools of EBM. In particular, the core of the development process is a systematic literature search followed by a review of the relevant peer-reviewed literature. Second, guidelines usually are focused on a clinical disorder (e.g., diabetes mellitus, stable angina pectoris) or a health care intervention (e.g., cancer screening). Third, the primary objective of guidelines is to improve the quality of medical care by identifying care practices that should be routinely implemented, based on high-quality evidence and high benefit-to-harm ratios for the interventions. Guidelines are intended to "assist" decision-making, not to define explicitly what decisions should be made in a particular situation, in part because guideline-level evidence alone is never sufficient for clinical decision-making (e.g., deciding whether to intubate and administer antibiotics for pneumonia in a terminally ill individual, in an individual with dementia, or in an otherwise healthy 30-year-old mother).

Guidelines are narrative documents constructed by expert panels whose composition often is determined by interested professional organizations. These panels vary in expertise and in the degree to which they represent all relevant stakeholders. The guideline documents consist of a series of specific management recommendations, a summary indication of the quantity and quality of evidence supporting each recommendation, an assessment of the benefit-to-harm ratio for the recommendation, and a narrative discussion of the recommendations. Many recommendations simply reflect the expert consensus of the guideline panel because literature-based evidence is insufficient or absent. A recent examination of this issue in cardiovascular guidelines showed that <15% of guideline recommendations were based on the highest level of clinical trial evidence, and this proportion had not improved in 10 years despite a substantial number of trials being conducted and published. The final step in guideline construction is peer review, followed by a final revision in response to the critiques provided.

Guidelines are closely tied to the process of quality improvement in medicine through their identification of evidence-based best practices. Such practices can be used as quality indicators. Examples include the proportion of acute MI patients who receive aspirin upon admission to a hospital and the proportion of heart failure patients with a depressed ejection fraction treated with an ACE inhibitor.

CONCLUSIONS

Thirty years after the introduction of the EBM movement, it is tempting to think that all the difficult decisions practitioners face have been or soon will be solved and digested into practice guidelines and computerized reminders. However, EBM provides practitioners with an ideal rather than a finished set of tools with which to manage patients. Moreover, even with such evidence, it is always worth remembering that the response to therapy of the "average" patient represented by the summary clinical trial outcomes may not be what can be expected for the specific patient sitting in front of a provider in the clinic or hospital. In addition, meta-analyses cannot generate evidence when there are no adequate randomized trials, and most of what clinicians confront in practice will never be thoroughly tested in a randomized trial. For the foreseeable future, excellent clinical reasoning skills and experience supplemented by well-designed quantitative tools and a keen appreciation for the role of individual patient preferences in their health care will continue to be of paramount importance in the practice of clinical medicine.

FURTHER READING

Croskerry P: A universal model of diagnostic reasoning. Acad Med 84:1022, 2009. [PubMed: 19638766]

Dhaliwal G, Detsky AS: The evolution of the master diagnostician. JAMA 310:579, 2013. [PubMed: 23942674]

Fanaroff AC et al: Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. JAMA 321:1069, 2019. [PubMed: 30874755]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Kahneman D: *Thinking Fast and Slow*. New York, Farrar, Straus and Giroux, 2013.

Fanaroff AC et al: Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. JAMA 321:1069, 2019. [PubMed: 30874755]

Hunink MGM et al: *Decision Making in Health and Medicine: Integrating Evidence and Values*, 2nd ed. Cambridge, Cambridge University Press, 2014.

Kahneman D: *Thinking Fast and Slow*. New York, Farrar, Straus and Giroux, 2013.

Kassirer JP et al: *Learning Clinical Reasoning*, 2nd ed. Baltimore, Lippincott Williams & Wilkins, 2009.

Mandelblatt JS et al: Collaborative modeling of the benefits and harms of associated with different U.S. breast cancer screening strategies. Ann Intern Med 164:215, 2016. [PubMed: 26756606]

Monteior S et al: The 3 faces of clinical reasoning: Epistemological explorations of disparate error reduction strategies. J Eval Clin Pract 24:666, 2018. [PubMed: 29532584]

Murthy VK et al: An inquiry into the early careers of master clinicians. J Grad Med Educ 10:500, 2018. [PubMed: 30386474]

Richards JB et al: Teaching clinical reasoning and critical thinking: From cognitive theory to practical application. Chest 158:1617, 2020. [PubMed: 32450242]

Royce CS et al: Teaching critical thinking: A case for instruction in cognitive biases to reduce diagnostic errors and improve patient safety. Acad Med 94:187, 2019. [PubMed: 30398993]

Saposnik G et al: Cognitive biases associated with medical decisions: A systematic review. BMC Med Inform Decis Mak 16:138, 2016. [PubMed: 27809908]

Schuwirth LWT et al: Assessment of clinical reasoning: three evolutions of thought. Diagnosis (Berl) 7:191, 2020. [PubMed: 32182208]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 4: Decision-Making in Clinical Medicine, Daniel B. Mark; John B. Wong
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 5: Precision Medicine and Clinical Care

The Editors

DISEASE NOSOLOGY AND PRECISION MEDICINE

Modern disease nosology arose in the late nineteenth century and represented a clear departure from the holistic, limited descriptions of disease dating to Galen. In this rubric, the definition of any disease is largely based on clinicopathologic observation. As the correlation between clinical signs and symptoms with pathoanatomy required autopsy material, diseases tended to be characterized by the end organ in which the primary syndrome was manifest and by late-stage presentations. Morgagni institutionalized this framework with the publication of *De Sedibus et Causis Morborum per Anatomen Indagatis* in 1761, in which he correlated the clinical features of patients with more than 600 autopsies at the University of Padua, demonstrating an anatomic basis for disease pathophysiology. Clinicopathologic observation served as the basis for inductive generalization coupled with the application of Occam's razor in which disease complexity was reduced to its simplest possible form. While this approach to defining human disease has held sway for over a century and facilitated the conquest of many diseases previously considered incurable, overly inclusive and simplified Oslerian diagnostics suffer from significant shortcomings. These include, but are not limited to, failure to distinguish the underlying etiology of different diseases with common pathophenotypes. For example, many different diseases can cause end-stage kidney disease or heart failure. Over time, the classification of neurodegenerative disorders or lymphomas, as well as many other diseases, is becoming more refined and precise as the underlying etiologies are identified. These distinctions are important for providing predictable prognostic information for individual patients with even highly prevalent diseases. Additionally, therapies may be ineffective owing to a lack of understanding of the often subtle molecular complexities of specific disease drivers.

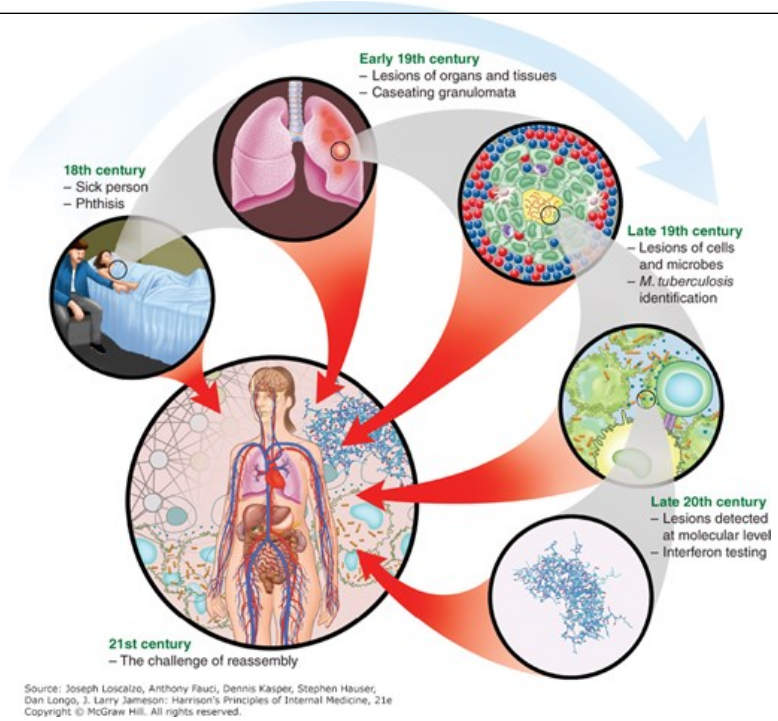
Beginning in the mid-twentieth century, the era of molecular medicine offered the idealized possibility of identifying the underlying molecular basis of every disease. Using a conventional reductionist paradigm, physician-scientists explored disease mechanism at ever-increasing molecular depth, seeking the single (or limited number of) molecular cause(s) of many human diseases. Yet, as effective as this now conventional scientific approach was at uncovering many disease mechanisms, the clinical manifestations of very few diseases could be explained on the basis of a single molecular mechanism. Even knowledge of the globin β chain mutation that causes sickle cell disease does not predict the many different manifestations of the disease (stroke syndrome, painful crises, and hemolytic crisis, among others). Clearly, the profession had expected too much from oversimplified reductionism and failed to take into consideration the extraordinary biologic variety and its accompanying molecular and genetic complexity that underpin both normal and pathologic diversity. The promise of the Human Genome Project provided new tools and approaches and unleashed efforts to identify a monogenic, oligogenic, or polygenic cause for every disease (allowing for environmental modulation). Yet, once again, disappointment reigned as the pool of genomes expanded without the expected revelations (aside from rare variants). The arc of progressive reductionism (as illustrated for tuberculosis in **Fig. 5-1**) in refining and explaining disease reached a humbling plateau, revealing the need for new approaches to understand better the etiology, manifestations, and progression of most diseases. The stage was set for a return to holism. However, in contrast to the holism of ancient physicians, we adopted one that is integrative, taking genomic context into account in all dimensions. In the course of elaborating this complex pathobiologic landscape, disease definition must become more precise and progressively more individualized, setting the stage for what we term *precision medicine*.

FIGURE 5-1

Arc of reductionism in medicine. *(Reproduced with permission from JA Greene, J Loscalzo: Putting the patient back together – social medicine, network medicine, and the limits of reductionism. N Engl J Med 377(25):2493-2499, 2017, Figure 1.)*

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS



Oversimplification of phenotype is a natural outgrowth of the observational scientific method. Categorizing individuals as falling into groups or clusters that are reasonably similar simplifies the task of the diagnostician and also facilitates the application of "specific" therapies more broadly. Biomedicine has been viewed as less quantitative and precise than other scientific disciplines, with biologic and pathobiologic diversity (biologic "noise") viewed as the norm. Thus, distilling such observational complexity to a fundamental group of symptoms or signs that are reasonably invariant across a group of sick individuals has served as the basis for the approach to disease and its treatment since the earliest days of medicine. This approach to diagnosis and therapy has remained in place into the twenty-first century, serving as the basis for the development of standard diagnostic tests and of broadly applied drug therapies. Targeting larger groups of patients is efficient when applied to large populations. As successful as this approach has been in advancing medical care, it is important to point out its limitations, which include significant predictive inaccuracies and sizeable segments of the disease population who do not respond to the most "effective" drugs (upward of 60% by some estimates). Clearly, a more nuanced approach to diagnosis and therapy is required to achieve better prognostic and therapeutic outcomes.

Turning first to phenotype, astute clinicians know full well the subtle and vivid differences in presentation that are often manifest among individuals with the same disease. In some cases, these differences in pathophenotype lead to new subclassifications of the disease, such as heart failure with preserved ejection fraction versus heart failure with reduced ejection fraction. Often, these relatively crude efforts at making diagnoses more precise are driven by new technologies or new ways of applying established technologies. In other cases, differences in pathophenotype are more subtle, not necessarily clinically apparent, and often driven by measures of endophenotype, such as distinctions among vasculitides facilitated by refinements in serologies or immunophenotyping. The impetus to create these subclasses of disease is largely determined by the need to improve prognosis and apply more precise and effective therapies. Based on these guiding principles, many experienced clinicians will argue—and rightly so—that they have been practicing personalized, precision medicine throughout their careers: they characterize each patient's illness in great detail, and choose therapies that respect and are guided by those individualized clinical and laboratory features, limited though they may be.

For many diseases, genomic variation, whether inherited or acquired, provides opportunities to refine diagnostic precision with even greater fidelity and predictive accuracy. For this reason, the field of precision medicine has now entered a new era that couples the molecular reductionism of the last century with an integrative, systems-level understanding of the basis for pathophenotype. Equally important, modern genomics has established that genomic context, sometimes referred to as modifier genes, is distinctive for each individual person; hence, understanding that context provides the insight necessary to predict how a primary disease driver or drivers may manifest a clinical pathophenotype—e.g., why some individuals with sickle cell anemia will develop stroke, while others will develop acute chest syndrome. *This concept that primary genetic and/or environmental drivers of a disease differentially affect disease expression based on an individual's unique genomic context serves as the ultimate basis for much of what we denote as precision medicine.*

To develop a precision medicine strategy for any disease, the clinician needs to be aware of two important, confounding principles. First, patients with

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

and antiphospholipid syndrome. Second, patients with the same basic disease can manifest very different pathophenotypes, i.e., *divergent*

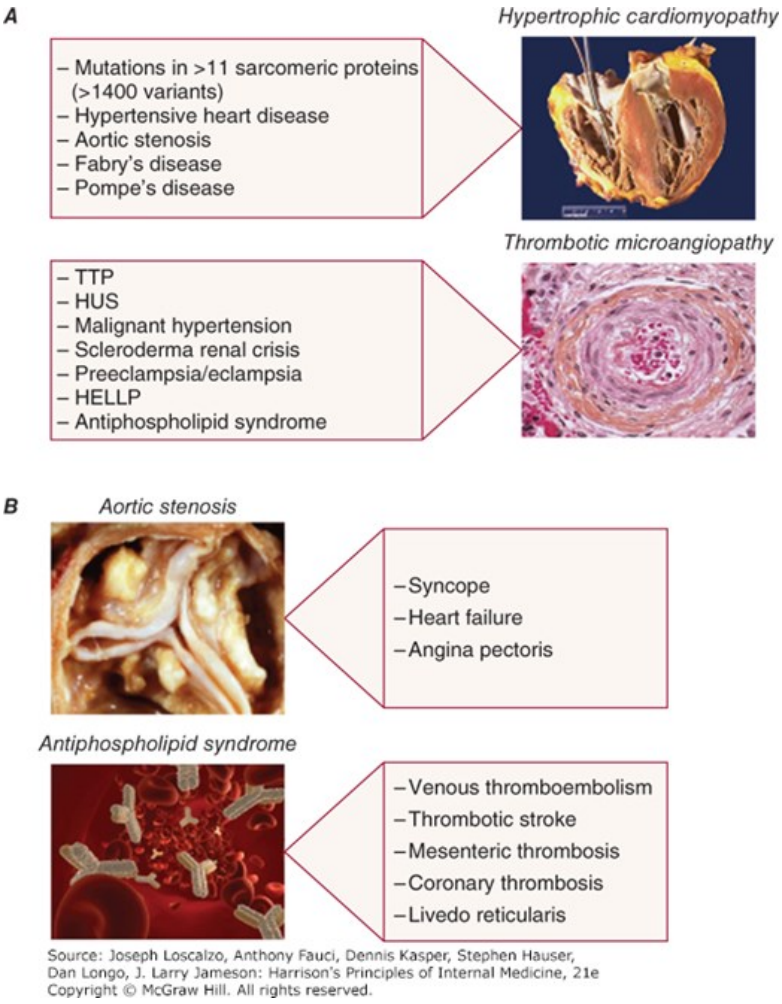
ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

To develop a precision medicine strategy for any disease, the clinician needs to be aware of two important, confounding principles. First, patients with different diseases can manifest similar pathophenotypes, i.e., *convergent phenotypes.* Examples of this principle include the hypertrophied myocardium found in hypertrophic cardiomyopathy, infiltrative cardiomyopathies, critical aortic stenosis, and untreated, long-standing hypertension; and the thrombotic microangiopathy found in malignant hypertension, scleroderma renal crisis, thrombotic thrombocytopenic purpura, eclampsia, and antiphospholipid syndrome. Second, patients with the same basic disease can manifest very different pathophenotypes, i.e., *divergent phenotypes* **(Chap. 466)**. Examples of this principle include the different clinical manifestations of cystic fibrosis or sickle cell disease and the incomplete penetrance of many common genetic diseases. These common presentations of different diseases and different presentations of the same disease are both a consequence of genomic context coupled with unique exposures over an individual's lifetime **(Fig. 5-2)**. Understanding the interplay among these many complex molecular determinants of disease expression is essential for the success of precision medicine.

FIGURE 5-2

Convergent and divergent phenotypes. Examples of the former (***A***) include hypertrophic cardiomyopathy and thrombotic microangiopathy, and examples of the latter, and (***B***) include aortic stenosis and antiphospholipid syndrome, each of which can have several distinct clinical presentations. HELLP, hemolysis, elevated liver enzymes, and a low platelet count; HUS, hemolytic-uremic syndrome; TTP, thrombotic thrombocytopenic purpura. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw-Hill Education. All rights reserved.)*



Given the complexity of the genomic and environmental context of an individual, one must ask the question: How precise do we need to be in order to practice effective precision medicine? Complete knowledge of a person's comprehensive genome (DNA, gene expression, mitochondrial function, proteome, metabolome, posttranslational modification of the proteome, and metagenome, among others) and quantitative assessments of environmental and social history are not possible to acquire; yet, this shortcoming does not render the general problem intractable. Owing to the fact that the molecular networks that govern phenotype are overdetermined (i.e., redundant) and that there are primary drivers of disease expression that

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

REQUIREMENTS FOR PRECISION MEDICINE

Copyright © McGraw Hill. All rights reserved.

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

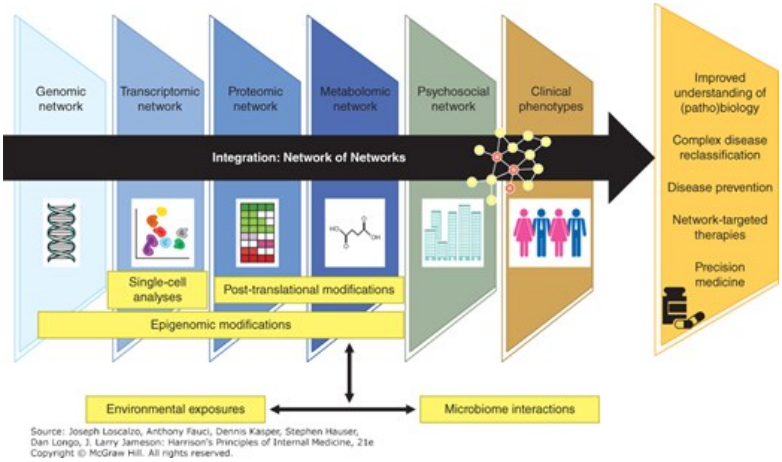
environmental and social history are not possible to acquire; yet, this shortcoming does not render the general problem intractable. Owing to the fact that the molecular networks that govern phenotype are overdetermined (i.e., redundant) and that there are primary drivers of disease expression that are modified in a weighted way by other genomic features of an individual, the practice of precision medicine can be realized without complete knowledge of all dimensions of the genome. Examples of how best to realize this strategy are discussed later in this chapter.

REQUIREMENTS FOR PRECISION MEDICINE

The essential elements of any precision medicine effort include phenotyping, endophenotyping (defining the characteristics of a disorder that are not readily observable), and genomic profiling **(Fig. 5-3)**. While subtle distinctions among individuals with the same disease are well known to clinicians, formalizing these nuanced differences is critical for achieving more precise phenotypes. Deep phenotyping requires a detailed history, including family history and environmental exposures, as well as relevant (physiologic) functional studies and imaging, including molecular imaging where appropriate. Biochemical, immunologic, and molecular tests of body fluids provide additional detail to the overall phenotype. Importantly, these objective laboratory tests together with functional studies compose an assessment of the endophenotype (or endotype) of an individual, refining the overall discriminant power of the evaluation. One additional concept that has gained traction in recent years is the notion of orthogonal phenotyping, i.e., assessing clinical, molecular, imaging, or functional (endo)phenotypes seemingly unrelated to the clinical presentation. These features further enhance the ability to distinguish (sub)phenotypes and derive from the fact that diseases can be subtly (subclinically) manifest in organ systems different from that in which the primary symptoms or signs are expressed. While some diseases are well known to affect multiple organ systems (e.g., systemic lupus erythematosus) and in many cases involvement of those many systems is assessed at initial diagnosis, such is not the case for most other diseases. As we begin to understand the differences in the organ-specific expression of genomic variants that drive or modify disease, it is becoming increasingly apparent that orthogonal—or more appropriately, unbiased comprehensive—phenotyping should become the norm.

FIGURE 5-3

Universe of precision medicine. The totality of precision medicine incorporates multidimensional biologic networks, the integration of which leads to a network of networks whose components interact with each other and with environmental exposures to yield a distinctive phenotype or pathophenotype. *(Reproduced with permission from LY-H Lee, J Loscalzo: Network medicine in pathobiology. Am J Pathol 189(7):1311-1326, 2019, Figure 3.)*



Genomic profiling must next be coupled to detailed phenotyping. The complex levels of genomic assessment continue to mature and include DNA sequencing (exomic, whole genome), gene expression (mRNA and protein expression), and metabolomics. In addition, the epigenome, the posttranslationally modified proteome, and the metagenome (the personal microbiome of an individual) are gaining traction as additional elements of comprehensive genomics **(Chap. 483)**. Not all of these genomic features are yet available for clinical laboratory testing, and those that are available are largely confined to blood testing. While DNA sequencing using whole blood would generally apply to any organ-based disease, gene expression, metabolomics, and epigenetics are often tissue specific. As tissue specimens cannot always or easily be obtained from the organ of interest, attempts at correlating whole-blood mRNA, protein, or metabolite profiles with those of the involved organ are critical for precise prognostics and therapeutic choices. In many cases, systemic consequences to an organ-specific disease (e.g., systemic inflammatory responses in individuals with atherosclerosis) can be ascertained and may provide useful prognostic information or therapeutic strategies. These biomarker signatures are the subject of ongoing discovery and have provided useful guidance toward improved diagnostic precision in many diseases. However, in many diseases, the correlations between these plasma or blood markers and organ-based diseases are weak, indicating a need to analyze each condition and each resulting signature under investigation is the conversion of induced pluripotent stem cells from a patient into a cell type of interest for gene expression or metabolomics study. As rational as this approach seems from first principles, it is important to note that gene expression patterns in these induced, differentiated cell

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

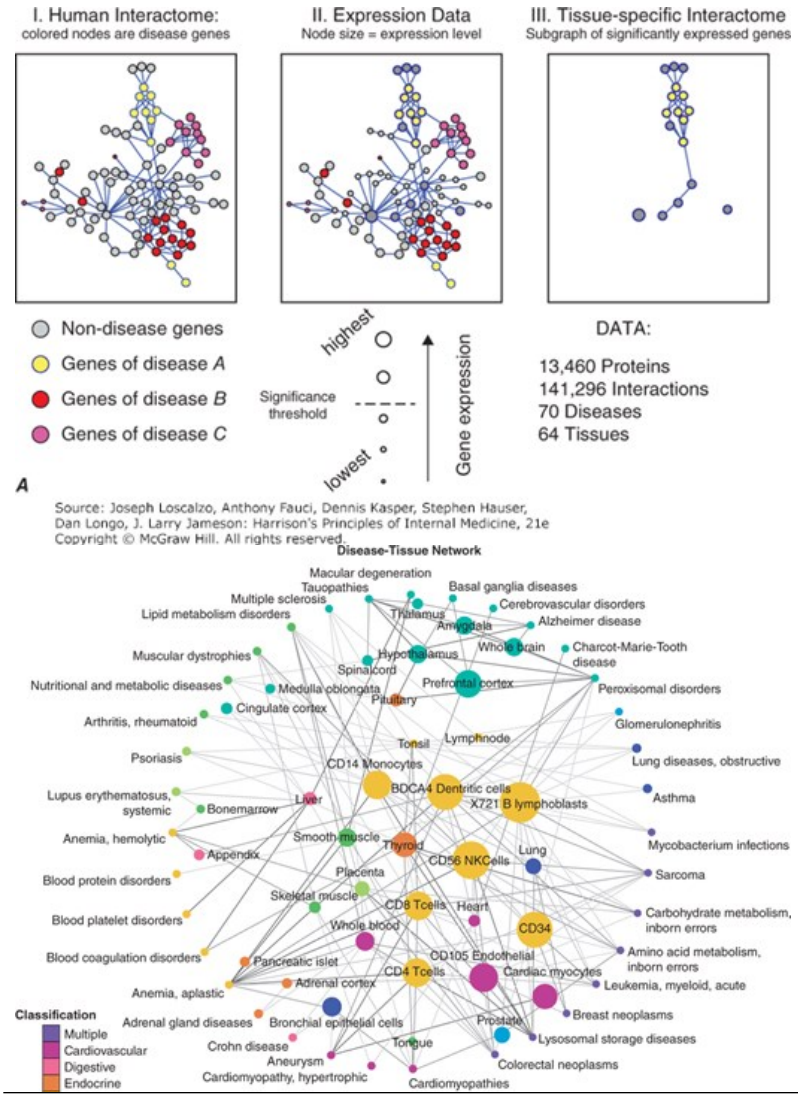
ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

discovery and have provided useful guidance toward improved diagnostic precision in many diseases. However, in many diseases, the correlations between these plasma or blood markers and organ-based diseases are weak, indicating a need to analyze each condition and each resulting signature before applying it to clinical decision-making. It is important to note that one of the key determinants of the functional consequences of a genetic variant believed to drive a disease phenotype is not simply its expression in a tissue of interest but, more importantly, the coexpression of protein binding partners in that same tissue comprising specific (dys)functional pathways that govern phenotype **(Fig. 5-4)**. An alternative strategy currently under investigation is the conversion of induced pluripotent stem cells from a patient into a cell type of interest for gene expression or metabolomics study. As rational as this approach seems from first principles, it is important to note that gene expression patterns in these induced, differentiated cell types are not completely consonant with their native counterparts, offering often limited additional information at potentially great additional expense.

FIGURE 5-4

Gene expression and phenotype. ***A.*** The human protein-protein interactome is constructed, and a specific disease module is identified (I); gene expression within this module is ascertained (II); and the tissue specificity of gene expression is determined (III). This analysis leads to a reduction of the total number of disease module genes that govern phenotype in a specific organ, which is a reflection of the specific pathway (or pathways) that is (or are) expressed in their functional entirety in that tissue. ***B.*** A disease-tissue bipartite network is constructed wherein specific tissues are placed within the circle and linked to diseases shown on the circumference. Nodes are colored according to tissue classification, the sizes of nodes are proportional to the total number of genes expressed in them, and the widths (shades) of the lines or edges correspond to the significance of the associations with specific diseases. *(From M Kitsak, A Sharma, J Menche et al: Tissue Specificity of Human Disease Module. Sci Rep 6, 35241, 2016, Figure 4.)*



Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

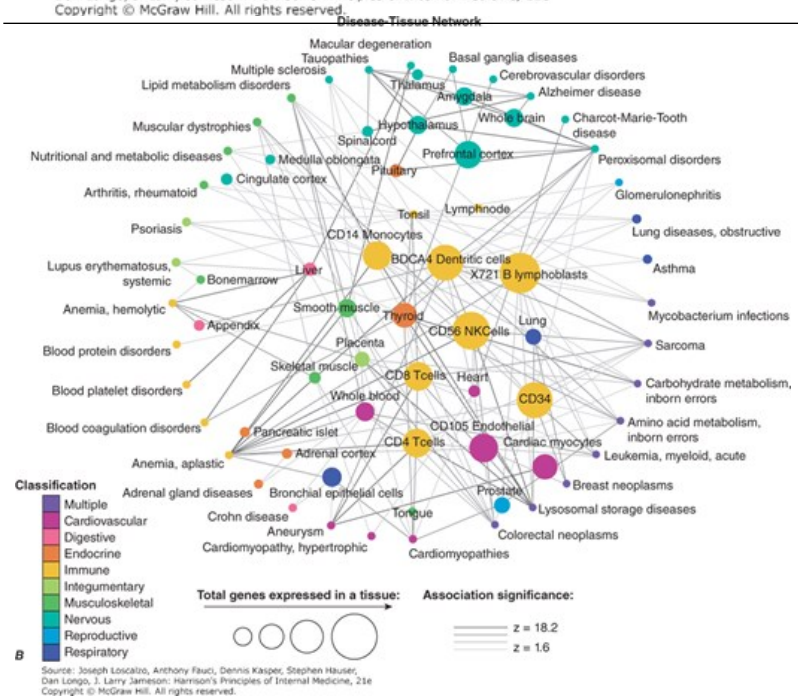
Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

While phenotype features of many chronic diseases are assessed over time, genomic features tend to be limited to single time point sampling. Time

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.



While phenotype features of many chronic diseases are assessed over time, genomic features tend to be limited to single time point sampling. Time trajectories are extremely informative in precision genotyping and phenotyping, with gene expression patterns and phenotypes changing over time in different ways among different patients with the same overarching phenotype. Cost, feasible sampling frequency, predictive power, and therapeutic choices will all drive the optimal strategy for the acquisition of timed samples in any given patient; however, with continued cost reduction in genomics technologies, this limitation may be progressively mitigated and clinical application may become a reality.

One important class of diseases that does not have most of these limitations in genomic profiling is cancer. Cancers can be (and are) sampled (biopsied) frequently to monitor temporal changes in the somatically mutating oncogenome and its consequences for the limited number of well-defined oncogenic driver pathways **(Chap. 68)**. A unique limitation of cancer in this regard, however, is that the frequency of somatic mutations over time (and, especially, with treatment) is great and the functional consequences of many of these mutations unknown. Equally important, assessment of single-cell mRNA sequencing patterns demonstrates great variability between apparently similar cells, challenging functional interpretation. Lastly, in solid tumors, stromal cells interact in a variety of ways (e.g., metabolically) with the associated malignant cells, and their gene expression signatures are also modified by the changing somatic mutational landscape of the primary malignancy. Thus, while much more information can be obtained over time in most cancer patients, the interpretation of these rich data sets continues to remain largely semi-empirical.

The possibility of identifying specific therapeutic targets remains a major goal of precision medicine. Doing so requires more than simple DNA sequencing and must include analysis of some level of gene expression, ideally in the involved organ(s). In addition to demonstrating the expression of a variant protein in the organ, one must ideally also demonstrate its functional consequences, which requires ascertaining the expression of binding partner proteins and the functional pathways they comprise. To achieve this goal, a variety of approaches have been tried, one of the most successful of which is the construction of the protein-protein interactome (the interactome), which is a comprehensive network map of the protein-protein interactions in a cell or organ of interest **(Chap. 486)**. This template provides information on the subnetworks that govern a disease phenotype (disease modules), which can be further individualized by incorporating individual variants and differentially expressed proteins that are patient specific. This type of analysis leads to the creation of an individual "reticulome" or reticulotype, which links the genotype to the phenotype of an individual **(Fig. 5-5)**. Using this approach, one can identify potential drug targets in a rational way or can even repurpose existing drugs by demonstrating the proximity of a known drug target to a disease module of interest **(Fig. 5-6)**. For example, in multicentric Castleman's disease, a disorder of unclear etiology, recognition that the PI3K/Akt/mTOR pathway is highly activated led to trials with an existing, approved drug, sirolimus. Precision medicine offers additional opportunities for optimizing the utilization of a drug by assessing the individualized pharmacogenomics of its disposition and metabolism, as demonstrated for the adverse consequences of variants in *TPMT* on azathioprine metabolism and in *CYP2C19* on clopidogrel metabolism **(Chap. 68)**.

FIGURE 5-5

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

integrative biologic network or reticulome derived from these multiomic analyses. These unique reticulotypes then serve as the basis for patient-specific, precision therapies. *(Adapted with permission from JA Leopold, J Loscalzo: Emerging Role of Precision Medicine in Cardiovascular Disease.*

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

FIGURE 5-5

Reticulotype. Patient-specific genotype-phenotype relationships by multiomic network structures are depicted for three individuals. Each individual's unique molecular perturbations (genetic variants, differentially expressed genes) are examined within the context of the subject's unique integrative biologic network or reticulome derived from these multiomic analyses. These unique reticulotypes then serve as the basis for patient-specific, precision therapies. *(Adapted with permission from JA Leopold, J Loscalzo: Emerging Role of Precision Medicine in Cardiovascular Disease. Circ Res 122:1302, 2018 and LYH Lee, J Loscalzo, Network Medicine in Pathobiology. Am J Pathol 189:1311, 2019, Figure 6.)*

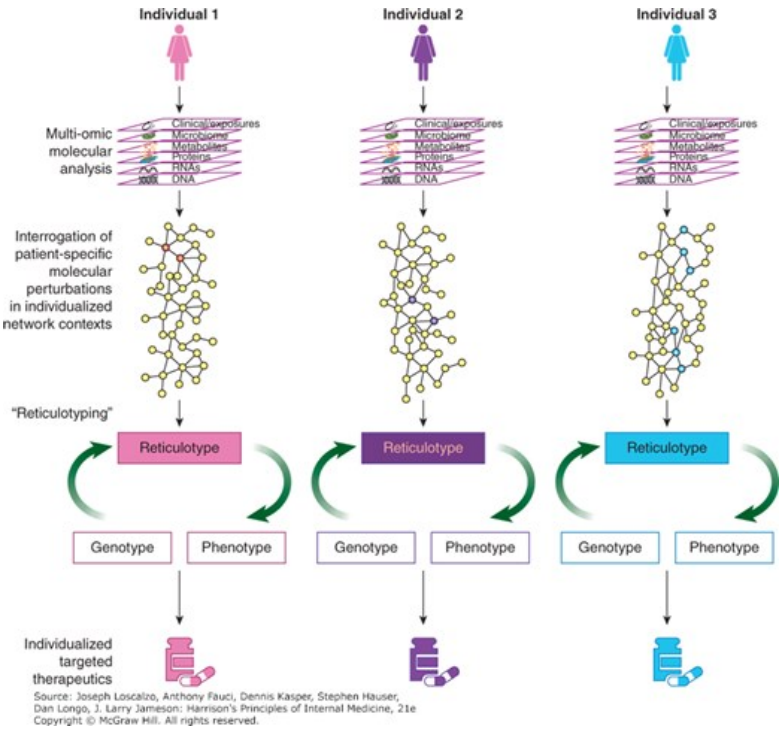
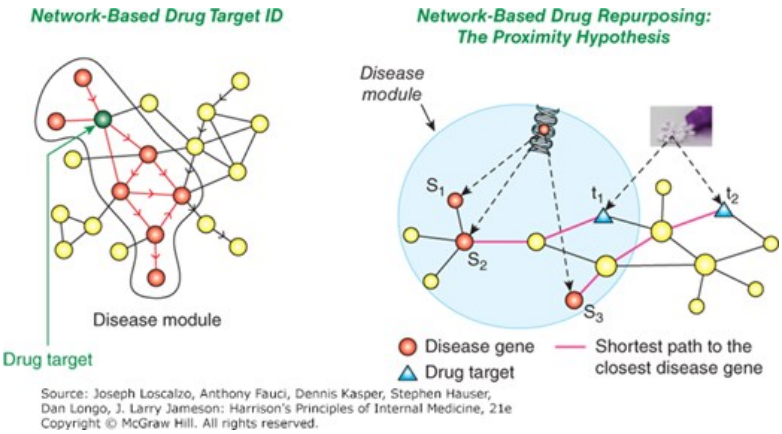


FIGURE 5-6

Network-based precision drug repurposing. *(Adapted from F Cheng et al: A genome-wide positioning systems network algorithm for in silico drug repurposing. Nat Commun 2019;10:3476, 2019.)*



EXAMPLES OF PRECISION MEDICINE APPLICATIONS

The field of precision medicine did not appear abruptly in medical history but, rather, evolved gradually as clinicians became more aware of differences among patients with the same disease. With the advent of modern genomics, in the ideal situation, these phenotype differences can now be mapped to genotype differences. Thus, we can consider precision medicine from the perspective of the pregenomic era and the postgenomic era.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

the evidence basis for their use, but tailored to the primary pathophysiologic phenotypes manifest in a patient, such as congestion, hypertension, and impaired contractility. These treatments were developed in the latter half of the last century based on empiric observation, reductionist experiments of

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

be mapped to genotype differences. Thus, we can consider precision medicine from the perspective of the pregenomic era and the postgenomic era.

Pregenomic precision medicine was applied to many diseases as therapeutic classes expanded for those disorders. A prime example of this approach is in the field of heart failure, where diuretics, digoxin, beta blockers, afterload-reducing agents, venodilators, renin-angiotensin-aldosterone inhibitors, and brain natriuretic peptide (nesiritide) are commonly used in some combination for most patients. The choice of agents is governed by the evidence basis for their use, but tailored to the primary pathophysiologic phenotypes manifest in a patient, such as congestion, hypertension, and impaired contractility. These treatments were developed in the latter half of the last century based on empiric observation, reductionist experiments of specific pathways believed to be involved in the pathophysiology, and clinical response in prospective trials. As phenotyping became more refined (e.g., echocardiographic assessments of ventricular function and tissue Doppler characterization of ventricular relaxation), the syndrome was subclassified into heart failure with reduced ejection fraction and heart failure with preserved ejection fraction, the latter of which does not respond well to any of the classes of therapeutic agents currently available. In the postgenomic era, ever more refined and detailed methods are under investigation to characterize pathophenotypes as well as genotypes, which may then be matched to the idealized combination of therapeutic classes of agents.

Pulmonary arterial hypertension is another disease for which definitive therapies straddle the pre- and postgenomic eras of precision medicine. Prior to the 1990s, there were no effective therapies for this highly morbid and lethal condition. With the advent of molecular and biochemical characterization of vascular abnormalities in individuals with established disease, however, therapies with agents that restored normal vascular function improved morbidity and mortality. These included calcium channel blockers, prostacyclin congeners, and endothelin receptor antagonists. As genomic characterization of the disease has progressed over the past two decades, there is increasing recognition of distinct genotypes that yield unique phenotypes **(Chap. 283)**, such as the demonstration of a primarily fibrotic endophenotype governed by the (oxidized) scaffold protein NEDD-9 and its aldosterone-dependent, TGF-β-independent enhancement of collagen III expression. This approach will continue to evolve as therapies become more effective (e.g., for perivascular fibrosis) and therapeutic choices better targeted to individual patients.

Precision genomics has also led to a new classification of the dementias, conditions previously thought to have a single cause with varied clinical expression. These disorders can now be categorized based on the genes and pathways involved and the site where aggregated proteins first form and then spread in the nervous system. For example, the varied clinical presentations of frontotemporal dementia, including progressive aphasia, behavioral disturbances, and dementia with amyotrophic lateral sclerosis, can now be linked to specific genotypes and susceptible cells **(Chap. 432)**. In prion diseases, the clinical phenotype is determined by specific germline mutations present in the prion protein **(Chap. 438)**. Discovery of autoantibodies against aquaporin-4 (AQP-4) and myelin oligodendrocyte glycoprotein (MOG) has allowed neuromyelitis optica, previously considered a multiple sclerosis–like disorder, to be classified as a separate entity requiring different treatment **(Chap. 445)**. Similarly, in myasthenia gravis, the identification of novel autoantibodies now permits stratification and a more finely tuned precision approach to therapy **(Chap. 448)**.

Precision medicine approaches to cancers have, of course, become the prime example of the opportunity that this strategy offers. In the pregenomic era, chemotherapy was widely used with variable success despite continued efforts to characterize the molecular features of the specific tumors and their semi-empiric responses to specific chemotherapeutic agents. As cancer genome sequencing evolved, however, it became apparent that there are a limited number of oncogenic pathways (<20) that are represented in the great majority of malignancies, without regard for the organ in which the disease was primarily manifest. These genomic signatures served as a template for precisely targeted therapies that have led to dramatic changes in response to treatment, including, for example, imatinib (and congeners) for Bcr-Abl tyrosine kinase activity in chronic myelogenous leukemia, erlotinib for *EGFR*-mutant non-small cell lung cancers, and ibrutinib for Bruton tyrosine kinase in chronic lymphocytic leukemia, among many others.

As exciting as these approaches have been, there are at least three primary challenges associated with precision therapeutics that are unique to cancer: (1) the mutational landscape continues to evolve as the disease progresses, and therapy often (if not invariably) leads to selection for resistant clones; (2) the likelihood that any cancer can be definitively cured by any single agent, no matter its exquisite precision, is quite limited, necessitating the development of rational polypharmaceutical approaches that take into account alternative pathways that achieve the same oncogenic goals as the primary targeted pathway, complicating drug development; and (3) there is marked genomic heterogeneity in many malignancies arguing that targeting a specific pathway—even with multiple drugs—may not ultimately succeed over the long term owing to the continued and heterogeneous evolution of the genomic landscape within a tumor within a patient. Despite these serious shortcomings, the application of progressively more refined and precisely targeted therapies used alone and in combination, such as with immune modulators, continues to offer great promise for the treatment of these diseases. In some ways, these approaches in cancer mirror earlier strategies in the treatment of infectious diseases in which the identification of the causative organism and its sensitivity to potential antimicrobials allows precision approaches to treatment. Combinatorial antimicrobial treatments represent an effective strategy to address acquired resistance. These diagnostic and therapeutic strategies can be applied without detailed knowledge of personalized responses to the infection or treatment (aside from serious adverse effects) with good outcomes in most cases. Yet, individuals do respond differently to specific infections and their treatments, possibly driven by different endophenotypes (e.g., different

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

THE FUTURE OF PRECISION MEDICINE

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

knowledge of personalized responses to the infection or treatment (aside from serious adverse effects) with good outcomes in most cases. Yet, individuals do respond differently to specific infections and their treatments, possibly driven by different endophenotypes (e.g., different inflammatory responses), suggesting that more precise knowledge of these precise mechanistic differences may yield improved prognosis and therapeutic approaches. As with cancer, immune modulation, particularly for immune exhaustion in chronic infections, represents a new frontier, again amenable to the personalized, precise analyses described above.

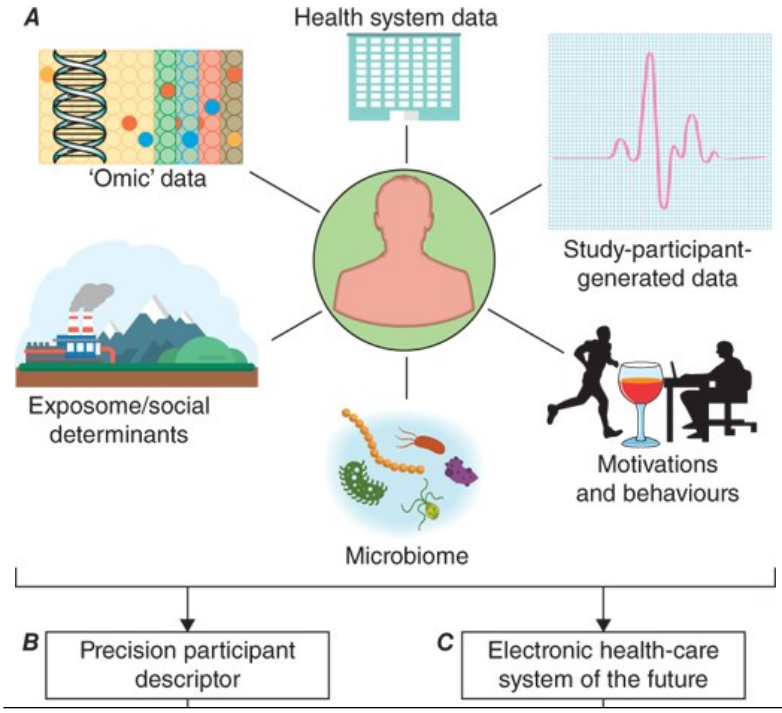
THE FUTURE OF PRECISION MEDICINE

Precision medicine clearly holds great promise for the future of the practice of medicine. For precision medicine to continue to evolve successfully, however, several requirements will need to be met. First, both deeply refined personal phenotypic data and genomic data are essential as the information with which precision analysis is performed. These data sets are quite large and require sufficient storage for analysis, especially for individuals in whom time trajectories are acquired (as should be the case for every person). Equally important, the analytical methods required to extract useful information from these data sets are evolving and themselves quite complex. While great progress has been made in genomics and biochemical testing, our ability to capture meaningful immunologic endophenotypes and environmental exposures is limited by comparison. Machine learning and artificial (auxiliary) intelligence methods will be essential for extracting optimal information from these data sets, which include not only pathways that can be uniquely targeted therapeutically but also individualized genomic or phenotypic signatures that are highly predictive of outcome, with or without therapy. Gathering sufficient information on the "normal" segments of the population is also required to ensure appropriate comparison data sets for optimal prediction.

Second, phenotyping must continue to expand and become dimensionally richer. The phenotypic features included in this data gathering must incorporate not only data relevant to the clinical presentation but also orthogonal phenotypic data that may yield useful information on disease trajectory or preclinical disease markers. Personal device data, environmental exposure history, social network interactions, and health system data will all be incorporated increasingly in defining phenotype and will require great efforts on the part of the medical informatics community to harmonize data sets, standardize data collection, and optimize/standardize data analysis **(Fig. 5-7)**.

FIGURE 5-7

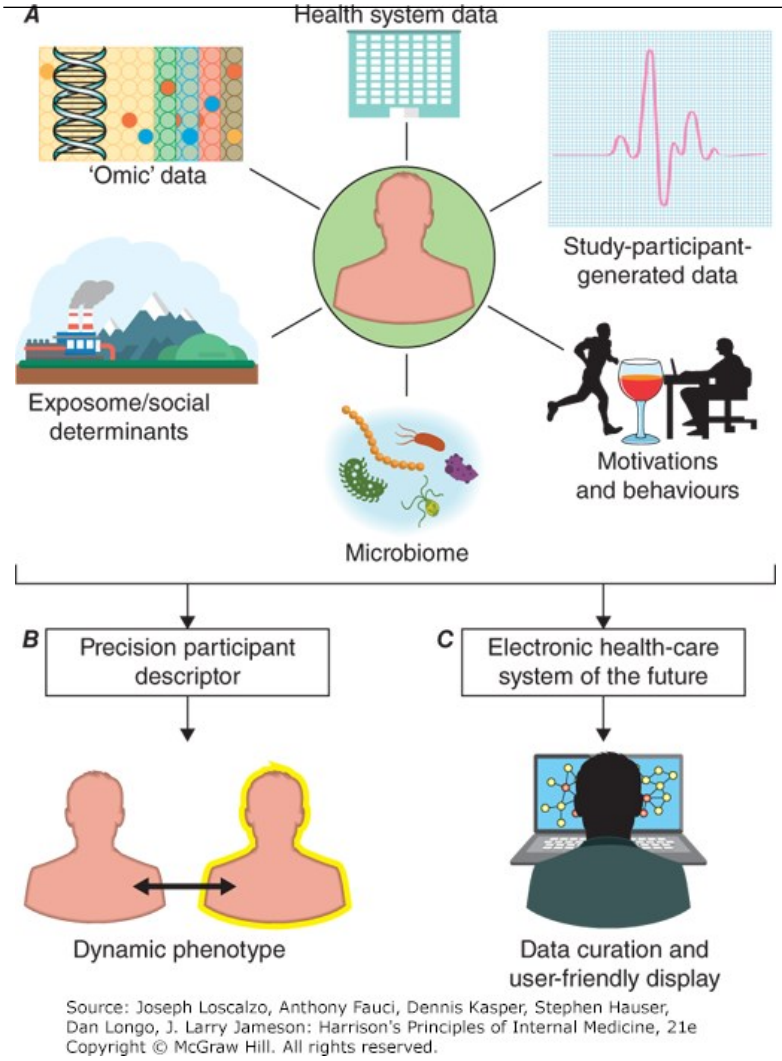
Big data in precision medicine. ***A.*** Six dimensions by which individuals may be characterized in the precision medicine era are described. ***B.*** The precision participant descriptor integrates the data from these six dimensions and varies over time. ***C.*** The electronic medical record increasingly must evolve to provide curated precision data in a user-friendly way. *(Reproduced with permission from EM Antman, J Loscalzo: Precision medicine in cardiology. Nat Rev Cardiol 13:591, 2016, Figure 2.)*



Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

cardiology. Nat Rev Cardiol 13:391, 2016, Figure 2.)



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Third, perhaps the greatest challenge to making precision medicine the standard approach to illness will be to determine the minimal data set required to predict outcome and response to therapy. Gathering data is comparatively simple; however, analyzing it to eliminate redundant information in these overdetermined biologic systems, weighting the determinants of an outcome, and using the data as phenomic/genomic signatures that are easier to collect than comprehensive, unbiased data sets are the ideal goals—a major challenge, but not insurmountable. Rapidly evolving machine learning and artificial intelligence strategies will also be essential for maximal success.

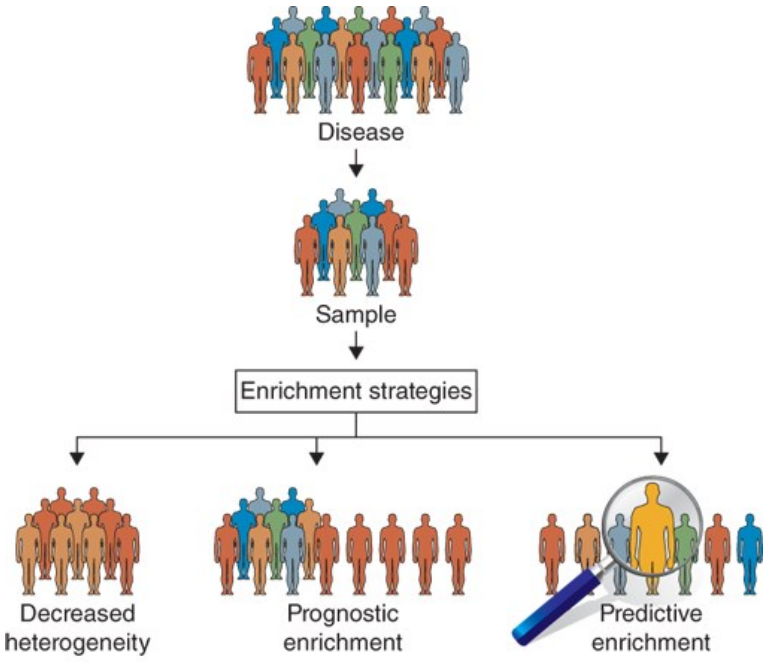
To return to the question of how precise precision medicine needs to be in order to be useful, please refer to **Fig. 5-8** where the approaches to clinical trial design meant to improve therapeutic signal are illustrated. Decreasing heterogeneity and enriching the study population will enhance the effect size, but these strategies are based on analyses of prior data sets that define those individuals who are more likely than not to respond to a therapy. By contrast, the notion of predictive enrichment follows from the information provided by a detailed, big data–driven analysis of individuals that explores phenotypic and genomic features used to predict response. These features need not be precisely met by each patient; however, they can be collated or clustered to define a reasonably sized cohort predicted to respond in a particular way within certain confidence bounds. In this way, the boundaries to the practice of precision medicine are imprecise strictly speaking, but sufficiently predictive to be practical from the perspectives of clinical care and cost-effectiveness.

FIGURE 5-8

The basis for precision medicine. The notion of precision medicine evolved, in part, from clinical trial design. From the entire population of patients with the disease of interest, a sample cohort of individuals is enrolled in the trial that ideally is representative of the entire distribution. Enrichment strategies developed to decrease heterogeneity or increase the representation of individuals with a high risk of observed outcomes (prognostic enrichment) facilitate trial conduct but do not necessarily improve precision in defining treatment response. The predictive enrichment

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Enrichment strategies developed to decrease heterogeneity or increase the representation of individuals with a high risk of observed outcomes (prognostic enrichment) facilitate trial conduct but do not necessarily improve precision in defining treatment response. The predictive enrichment strategy utilizes both trial participant characteristics and data from experiments conducted before or during (adaptive design) the trial to improve the prediction of who is likely to have a more pronounced response to the treatment under study. *(Reproduced with permission from EM Antman, J Loscalzo: Precision medicine in cardiology. Nat Rev Cardiol 13:591, 2016, Figure 3.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FURTHER READING

Antman EM, Loscalzo J: Precision medicine in cardiology. Nat Rev Cardiol 13:591, 2016. [PubMed: 27356875]

Cheng F et al: Comprehensive characterization of protein-protein interactions perturbed by disease mutations. Nat Genet 53:342, 2021. [PubMed: 33558758]

Cheng F et al: A genome-wide positioning systems network algorithm for in silico drug repurposing. Nat Commun 10:3476, 2019. [PubMed: 31375661]

Greene JA, Loscalzo J: Putting the patient back together—Social medicine, network medicine, and the limits of reductionism. N Engl J Med 377:2493, 2017. [PubMed: 29262277]

Kitsak M et al: Tissue specificity of human disease module. Sci Rep 6:35241, 2016. [PubMed: 27748412]

Lee LY, Loscalzo J: Network medicine in pathobiology. Am J Pathol 189:1311, 2019. [PubMed: 31014954]

Leopold JA et al: The application of big data to cardiovascular disease: Paths to precision medicine. J Clin Invest 130:29, 2020. [PubMed: 31895052]

Loscalzo J et al: Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol 3:124, 2007. [PubMed: 17625512]

Maron BA et al: Individualized interactomes for network-based precision medicine in hypertrophic cardiomyopathy with implications for other clinical pathophenotypes. Nat Commun 12:873, 2021. [PubMed: 33558530]

Menche J et al: Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science 347:1257601, 2015. [PubMed: 29899023]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Menche J et al: Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science 347:1257601, 2015.
[PubMed: 25700523]

Samokhin AO et al: NEDD9 targets *COL3A1* to promote endothelial fibrosis and pulmonary arterial hypertension. Sci Transl Med 10:eaap7294, 2018.
[PubMed: 29899023]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 5: Precision Medicine and Clinical Care, The Editors
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 6: Screening and Prevention of Disease

Katrina A. Armstrong; Gary J. Martin

INTRODUCTION

A primary goal of health care is to prevent disease or detect it early enough that intervention will be more effective. Tremendous progress has been made toward this goal over the past 50 years. Screening tests are available for many common diseases and encompass biochemical (e.g., cholesterol, glucose), physiologic (e.g., blood pressure, growth curves), radiologic (e.g., mammogram, bone densitometry), and cytologic (e.g., Pap smear) approaches. Effective preventive interventions have resulted in dramatic declines in mortality from many diseases, particularly infections. Preventive interventions include counseling about risk behaviors, vaccinations, medications, and, in some relatively uncommon settings, surgery. Preventive services (including screening tests, preventive interventions, and counseling) are different than other medical interventions because they are proactively administered to healthy individuals instead of in response to a symptom, sign, or diagnosis. Thus, the decision to recommend a screening test or preventive intervention requires a particularly high bar of evidence that testing and intervention are both practical and effective.



Because population-based screening and prevention strategies must be extremely low risk to have an acceptable benefit-to-harm ratio, the ability to target individuals who are more likely to develop disease could enable the application of a wider set of potential approaches and increase efficiency. Currently, there are many types of data that can predict disease incidence in an asymptomatic individual. Germline genomic data have received the most attention to date, at least in part because mutations in high-penetrance genes have clear implications for preventive care **(Chap. 467)**. Women with mutations in either *BRCA1* or *BRCA2*, the two major breast cancer susceptibility genes identified to date, have a markedly increased risk (five- to twentyfold) of breast and ovarian cancer. Screening and prevention recommendations include prophylactic oophorectomy and breast magnetic resonance imaging (MRI), both of which are considered to incur too much harm for women at average cancer risk. Some women with *BRCA* mutations opt for prophylactic mastectomy to dramatically reduce their breast cancer risk. Although the proportion of common disease explained by high-penetrance genes appears to be relatively small (5–10% of most diseases), mutations in rare, moderate-penetrance genes, and variants in low-penetrance genes, also contribute to the prediction of disease risk. Most recently, polygenic risk scores combining information about variants across hundreds of genes are being evaluated for identifying individuals at high risk of coronary heart disease and other conditions. The advent of affordable whole exome/whole genome sequencing is likely to speed the dissemination of these tests into clinical practice and may transform the delivery of preventive care.

Other forms of “omic” data also have the potential to provide important predictive information. Proteomics and metabolomics can provide insight into gene function, but it has proven challenging to develop reliable, predictive measures using these platforms. More recently, it has become possible to measure the presence of mutations in DNA circulating in the bloodstream and in stool, with early promising evidence that these assays can be used to detect cancer before existing screening tests.

In addition to “omic” data, imaging data are increasingly being integrated into risk-stratified prevention approaches as evidence grows about the predictive ability of these data. For example, coronary computed tomography (CT) scans are used in many preventive cardiology programs to inform decisions about beginning statin therapy when there is conflicting or uncertain information from other risk assessment approaches. Of course, these data may also be helpful in predicting the risk of harms from screening or prevention, such as the risk of a false-positive mammogram.

In addition to advances in risk prediction, there are several other reasons that screening and prevention are likely to gain importance in medical care in the near term. New imaging modalities are being developed that promise to detect changes at the cellular and subcellular levels, greatly increasing the probability that early detection improves outcomes. The rapidly growing understanding of the biologic pathways underlying initiation and progression of many common diseases has the potential to transform the development of preventive interventions, including chemoprevention. Furthermore, screening and prevention offer the promise of both improving health and sparing the costs of disease treatment, an issue that will continue to gain importance as long as health care costs in the United States remain a concern to patients, government agencies, and insurers.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

BASIC PRINCIPLES OF SCREENING

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

importance as long as health care costs in the United States remain a concern to patients, government agencies, and insurers.

This chapter will review the basic principles of screening and prevention in the primary care setting. Recommendations for specific disorders such as cardiovascular disease, diabetes, and cancer are provided in the chapters dedicated to those topics.

BASIC PRINCIPLES OF SCREENING

The basic principles of screening populations for disease were published by the World Health Organization in 1968 **(Table 6-1)**.

TABLE 6-1

Principles of Screening

The condition should be an important health problem.
There should be a treatment for the condition.
Facilities for diagnosis and treatment should be available.
There should be a latent stage of the disease.
There should be a test or examination for the condition.
The test should be acceptable to the population.
The natural history of the disease should be adequately understood.
There should be an agreed policy on whom to treat.
The cost of finding a case should be balanced in relation to overall medical expenditure.

In general, screening is most effective when applied to relatively common disorders that carry a large disease burden **(Table 6-2)**. The five leading causes of mortality in the United States are heart diseases, malignant neoplasms, chronic obstructive pulmonary disease, accidents, and cerebrovascular diseases. Thus, many screening strategies are targeted at these conditions. From a global health perspective, these conditions are priorities, but malaria, malnutrition, AIDS, tuberculosis, and violence also carry a heavy disease burden **(Chap. 472)**.

TABLE 6-2

Lifetime Cumulative Risk

Breast cancer for women	10%
Colon cancer	6%
Cervical cancer for women[a]	2%
Domestic violence for women	Up to 15%
Hip fracture for white women	16%

[a]Assuming an unscreened population.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

development of "see and treat" approaches such as those currently used for cervical cancer screening in some countries. A long latent or preclinical phase where early treatment increases the chance of cure is a hallmark of many cancers; for example, polypectomy prevents progression to colon

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Having an effective treatment for early disease has proven challenging for some common diseases. For example, although Alzheimer's disease is the sixth leading cause of death in the United States, there are no curative treatments and no evidence that early treatment improves outcomes. Lack of facilities for diagnosis and treatment is a particular challenge for developing countries and may change screening strategies, including the development of "see and treat" approaches such as those currently used for cervical cancer screening in some countries. A long latent or preclinical phase where early treatment increases the chance of cure is a hallmark of many cancers; for example, polypectomy prevents progression to colon cancer. Similarly, early identification of hypertension or hyperlipidemia allows therapeutic interventions that reduce the long-term risk of cardiovascular or cerebrovascular events. In contrast, lung cancer screening has historically proven more challenging because most tumors are not curable by the time they can be detected on a chest x-ray. However, the length of the preclinical phase also depends on the level of resolution of the screening test, and this situation changed with the development of chest CT. Low-dose chest CT scanning can detect tumors earlier and has been demonstrated to reduce lung cancer mortality by 20% in individuals who had at least a 30-pack-year history of smoking. The short interval between the ability to detect disease on a screening test and the development of incurable disease also contributes to the limited effectiveness of mammography screening in reducing deaths from some forms of breast cancer. At the other end of the spectrum, the early detection of prostate cancer may not lead to a difference in the mortality rate because the disease is often indolent and competing morbidities, such as coronary artery disease, may ultimately cause mortality **(Chap. 70)**. This uncertainty about the natural history is also reflected in the controversy about treatment of prostate cancer, further contributing to the challenge of screening in this disease. Finally, screening programs can incur significant economic costs that must be considered in the context of the available resources and alternative strategies for improving health outcomes.

METHODS OF MEASURING HEALTH BENEFITS

Because screening and preventive interventions are recommended to asymptomatic individuals, they are held to a high standard for demonstrating a favorable risk-benefit ratio before implementation. In general, the principles of evidence-based medicine apply to demonstrating the efficacy of screening tests and preventive interventions, where randomized controlled trials (RCTs) with mortality outcomes are the gold standard. However, because RCTs are often not feasible, observational studies, such as case-control designs, have been used to assess the effectiveness of some interventions such as colonoscopy for colorectal cancer screening. For some strategies, such as Pap smear screening for cervical cancer, the only data available are ecologic data demonstrating dramatic declines in mortality.

Irrespective of the study design used to assess the effectiveness of screening, it is critical that disease incidence or mortality is the primary endpoint rather than length of disease survival. This is important because lead time bias and length time bias can create the appearance of an improvement in disease survival from a screening test when there is no actual effect. Lead time bias occurs because screening identifies a case before it would have presented clinically, thereby creating the perception that a patient lived longer after diagnosis simply by moving the date of diagnosis earlier rather than the date of death later. Length time bias occurs because screening is more likely to identify slowly progressive disease than rapidly progressive disease. Thus, within a fixed period of time, a screened population will have a greater proportion of these slowly progressive cases and will appear to have better disease survival than an unscreened population.

A variety of endpoints are used to assess the potential gain from screening and preventive interventions.

1. *The absolute and relative impact of screening on disease incidence or mortality*. The absolute difference in disease incidence or mortality between a screened and nonscreened group allows the comparison of size of the benefit across preventive services. A meta-analysis of Swedish mammography trials (ages 40–70) found that ~1.2 fewer women per 1000 would die from breast cancer if they were screened over a 12-year period. By comparison, at least ~3 lives per 1000 would be saved from colon cancer in a population (aged 50–75) screened with annual fecal occult blood testing (FOBT) over a 13-year period, and an estimated 20–24 lives per 1000 would be saved over the entire 25-year period. Based on this analysis, colon cancer screening may actually save more women's lives than does mammography. However, the relative impact of FOBT (30% reduction in colon cancer death) is similar to the relative impact of mammography (14–32% reduction in breast cancer death), emphasizing the importance of both relative and absolute comparisons.

2. *The number of subjects screened to prevent disease or death in one individual*. The inverse of the absolute difference in mortality is the number of subjects who would need to be screened or receive a preventive intervention to prevent one death. For example, 731 women aged 65–69 would need to be screened by dual-energy x-ray absorptiometry (DEXA) (and treated appropriately) to prevent one hip fracture from osteoporosis.

3. *Increase in average life expectancy for a population*. Predicted increases in life expectancy for various screening and preventive interventions are listed in **Table 6-3**. It should be noted, however, that the increase in life expectancy is an average that applies to a population, not to an individual.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

reasonable goal for a population-based screening or prevention strategy.

TABLE 6-3

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

subjects who would need to be screened or receive a preventive intervention to prevent one death. For example, 721 women aged 65–69 would

listed in **Table 6-3**. It should be noted, however, that the increase in life expectancy is an average that applies to a population, not to an individual. In reality, the vast majority of the population does not derive any benefit from a screening test. A small subset of patients, however, will benefit greatly. For example, Pap smears do not benefit the 98% of women who never develop cancer of the cervix. However, for the 2% who would have developed cervical cancer, Pap smears may add as much as 25 years to their lives. Some studies suggest that a 1-month gain of life expectancy is a reasonable goal for a population-based screening or prevention strategy.

TABLE 6-3

Estimated Average Increase in Life Expectancy for a Population

SCREENING OR PREVENTIVE INTERVENTION	AVERAGE INCREASE
Mammography:	
Women, 40–50 years	0–5 days
Women, 50–70 years	1 month
Pap smears, age 18–65	2–3 months
Getting a 35-year-old smoker to quit	3–5 years
Beginning regular exercise for a 40-year-old man (30 min, 3 times a week)	9 months–2 years

ASSESSING THE HARMS OF SCREENING AND PREVENTION

Just as with most aspects of medical care, screening and preventive interventions also incur the possibility of adverse outcomes. These adverse outcomes include side effects from preventive medications and vaccinations, false-positive screening tests, overdiagnosis of disease from screening tests, anxiety, radiation exposure from some screening tests, and discomfort from some interventions and screening tests. The risk of side effects from preventive medications is analogous to the use of medications in therapeutic settings and is considered in the U.S. Food and Drug Administration (FDA) approval process. Side effects from currently recommended vaccinations are primarily limited to discomfort and minor immune reactions. However, the concern about associations between vaccinations and serious adverse outcomes continues to limit the acceptance of many vaccinations despite the lack of data supporting the causal nature of these associations.

The possibility of a false-positive test occurs with nearly all screening tests, although the definition of what constitutes a false-positive result often varies across settings. For some tests such as screening mammography and screening chest CT, a false-positive result occurs when an abnormality is identified that is not malignant, requiring either a biopsy diagnosis or short-term follow-up. For other tests such as Pap smears, a false-positive result occurs because the test identifies a wide range of potentially premalignant states, only a small percentage of which would ever progress to an invasive cancer. This risk is closely tied to the risk of overdiagnosis in which the screening test identifies disease that would not have presented clinically in the patient's lifetime. Assessing the degree of overdiagnosis from a screening test is very difficult given the need for long-term follow-up of an unscreened population to determine the true incidence of disease over time. Recent estimates suggest that as much as 15–40% of breast cancers identified by mammography screening and 15–37% of prostate cancers identified by prostate-specific antigen testing may never have presented clinically. Screening tests also have the potential to create unwarranted anxiety, particularly in conjunction with false-positive findings. Although multiple studies have documented increased anxiety through the screening process, there are few data suggesting this anxiety has long-term adverse consequences, including subsequent screening behavior. Screening tests that involve radiation (e.g., mammography, chest CT) add to the cumulative radiation exposure for the screened individual. The absolute amount of radiation is very small from any of these tests, but the overall impact of repeated exposure from multiple sources is still being determined. Some preventive interventions (e.g., vaccinations) and screening tests (e.g., mammography) may lead to discomfort at the time of administration, but again, there is little evidence of long-term adverse consequences.

WEIGHING THE BENEFITS AND HARMS

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

saved, with adjustment for the quality of life impact of different interventions and disease states (i.e., quality-adjusted life-year). Typically, strategies that cost $50,000–100,000 per quality-adjusted year of life saved are considered "cost-effective" **(Chap. 4)**

WEIGHING THE BENEFITS AND HARMS

The decision to implement a population-based screening and prevention strategy requires weighing the benefits and harms, including the economic impact of the strategy. The costs include not only the expense of the intervention but also time away from work, downstream costs from false-positive results, "incidentalomas" or adverse events, and other potential harms. Cost-effectiveness is typically assessed by calculating the cost per year of life saved, with adjustment for the quality of life impact of different interventions and disease states (i.e., quality-adjusted life-year). Typically, strategies that cost $50,000–100,000 per quality-adjusted year of life saved are considered "cost-effective" **(Chap. 4)**.

The U.S. Preventive Services Task Force (USPSTF) is an independent panel of experts in preventive care that provides evidence-based recommendations for screening and preventive strategies based on an assessment of the benefit-to-harm ratio **(Tables 6-4 and 6-5)**. Because there are multiple advisory organizations providing recommendations for preventive services, the agreement among the organizations varies across the different services. For example, all advisory groups support screening for hyperlipidemia and colorectal cancer, whereas consensus is lower for breast cancer screening among women in their forties and for prostate cancer screening. Because the guidelines are only updated periodically, differences across advisory organizations may also reflect the data that were available when the guideline was issued.

TABLE 6-4

Screening Tests Recommended by the U.S. Preventive Services Task Force for Average-Risk Adults

DISEASE	TEST	POPULATION	FREQUENCY	CHAPTER
Abdominal aortic aneurysm	Ultrasound	Men 65–75 who have ever smoked	Once	
Alcohol misuse	Alcohol Use Disorders Identification Test	All adults	Unknown	453
Breast cancer	Mammography with or without clinical breast examination	Women 50–75	Every 2 years	
Cervical cancer	Pap smear	Women 21–65	Every 3 years	70
	Pap smear and/or HPV testing	Women 30–65	Every 5 years if HPV negative	
Chlamydia/gonorrhea	Nucleic acid amplification test on urine or cervical swab	Sexually active women <25	Unknown	189
Colorectal cancer	Fecal occult blood testing	45–75	Every year	70, 81
	Fecal immunochemical-DNA	45–75	Every 1–3 years	
	Sigmoidoscopy	45–75	Every 5 years	
	Colonoscopy (or occult blood testing combined with sigmoidoscopy)	45–75	Every 10 years	
Depression	Screening questions	All adults	Periodically	
Diabetes	Fasting blood glucose or HgbA1c	Adults overweight, obese, or with hypertension	Every 3 years	403
HIV	Reactive immunoassay or rapid HIV followed by confirmatory test	15–65	At least once	

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Diabetes	Fasting blood glucose or HgbA1c	Adults overweight, obese, or with hypertension	Every 3 years	403
Hepatitis C	Anti-HCV antibody followed by confirmatory PCR	18–79	Once	
HIV	Reactive immunoassay or rapid HIV followed by confirmatory test	15–65	At least once	
Hyperlipidemia	Cholesterol	40–75	Unknown	407
Hypertension	Blood pressure	All adults	Periodically	277
Intimate partner violence	Screening questions	Women of childbearing age	Unknown	
Lung cancer	Low-dose computed tomography	Adults 50 to 80 years who have a 20 pack-year smoking history and currently smoke or have quit within the past 15 years	Yearly	
Obesity	Body mass index	All adults	Unknown	
Osteoporosis	DEXA	Women >65 or >60 with risk factors	Unknown	411

Abbreviations: DEXA, dual-energy x-ray absorptiometry; HCV, hepatitis C virus; HPV, human papillomavirus; PCR, polymerase chain reaction.

Source: Adapted from the U.S. Preventive Services Task Force 2017. *www.uspreventiveservicestaskforce.org/Page/Name/uspstf-a-and-b-recommendations/.*

TABLE 6-5

Preventive Interventions Recommended for Average-Risk Adults

INTERVENTION	DISEASE	POPULATION	FREQUENCY	CHAPTER
Adult immunization				123, 124
Tetanus-diphtheria		>18	Every 10 years	
Varicella		Susceptibles only, >18	Two doses	
Measles-mumps-rubella		Women, childbearing age	One dose	
Pneumococcal		>64	13 followed by 23 valent	
Influenza		>18	Yearly	
Human papillomavirus		Up to age 27	If not done prior	
Zoster		>60	Once	
Chemoprevention				
Folic acid	Neural tube defects in baby	Women planning or capable of pregnancy		

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Source: Adapted from the U.S. Preventive Services Task Force 2017. www.uspreventiveservicestaskforce.org/Page/Name/uspstf-a-and-b-recommendations/.

TABLE 6-5

Preventive Interventions Recommended for Average-Risk Adults

INTERVENTION	DISEASE	POPULATION	FREQUENCY	CHAPTER
Adult immunization				123, 124
Tetanus-diphtheria		>18	Every 10 years	
Varicella		Susceptibles only, >18	Two doses	
Measles-mumps-rubella		Women, childbearing age	One dose	
Pneumococcal		>64	13 followed by 23 valent	
Influenza		>18	Yearly	
Human papillomavirus		Up to age 27	If not done prior	
Zoster		>60	Once	
Chemoprevention				
Aspirin	Cardiovascular disease	Aged 50–59 years with a ≥10% 10-year cardiovascular disease risk (bleeding risk may = benefit for some groups)		
Folic acid	Neural tube defects in baby	Women planning or capable of pregnancy		
Tamoxifen/raloxifene	Breast cancer	Women at high risk for breast cancer		
Vitamin D	Fracture/falls	>64 at increased risk for falls		

For many screening tests and preventive interventions, the balance of benefits and harms may be uncertain for the average-risk population but more favorable for individuals at higher risk for disease. Although age is the most commonly used risk factor for determining screening and prevention recommendations, the USPSTF also recommends some screening tests in populations based upon the presence of other risk factors for the disease. In addition, being at increased risk for the disease often supports initiating screening at an earlier age than that recommended for the average-risk population. For example, when there is a significant family history of colon cancer, it is prudent to initiate screening 10 years before the age at which the youngest family member was diagnosed with cancer.

Although informed consent is important for all aspects of medical care, shared decision-making may be a particularly important approach to decisions about preventive services when the benefit-to-harm ratio is uncertain for a specific population. For example, many expert groups, including the American Cancer Society, recommend an individualized discussion about prostate cancer screening, because the decision-making process is complex and relies heavily on personal issues. Some men may decline screening, whereas others may be more willing to accept the risks of an early detection strategy. Recent analysis suggests that many men may be better off not screening for prostate cancer because watchful waiting was the preferred strategy when quality-adjusted life-years were considered. Another example of shared decision-making involves the choice of techniques for colon cancer screening **(Chap. 70)**. In controlled studies, the use of annual FOBT reduces colon cancer deaths by 15–30%. Flexible sigmoidoscopy reduces

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Thus, although one patient may prefer the ease of preparation, less time disruption, and the lower risk of flexible sigmoidoscopy, others may prefer the sedation, thoroughness, and time interval of colonoscopy.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

cancer screening **(Chap. 70)**. In controlled studies, the use of annual FOBT reduces colon cancer deaths by 15–30%. Flexible sigmoidoscopy reduces colon cancer deaths by ~40–60%. Colonoscopy appears to offer a greater benefit than flexible sigmoidoscopy with a reduction in risk of ~70%, but its use incurs additional costs and risks. These screening procedures have not been compared directly in the same population, but models suggest that appropriate frequencies of each technique may be associated with similar numbers of lives saved and cost to society per life saved ($10,000–25,000). Thus, although one patient may prefer the ease of preparation, less time disruption, and the lower risk of flexible sigmoidoscopy, others may prefer the sedation, thoroughness, and time interval of colonoscopy.

COUNSELING ON HEALTHY BEHAVIORS

In considering the impact of preventive services, it is important to recognize that tobacco and alcohol use, diet, and exercise constitute the vast majority of factors that influence preventable deaths in developed countries. Perhaps the single greatest preventive health care measure is to help patients quit smoking **(Chap. 454)**. However, efforts in these areas frequently require behavior changes (e.g., weight loss, exercise) or the management of addictive conditions (e.g., tobacco and alcohol use) that are often recalcitrant to intervention. Although these are challenging problems, evidence strongly supports the role of counseling by health care providers **(Table 6-6)** in effecting health behavior change. Educational campaigns, public policy changes, and community-based interventions have also proven to be important parts of a strategy for addressing these factors in some settings. Although the USPSTF found that the evidence was conclusive to recommend a relatively small set of counseling activities, counseling in areas such as physical activity and injury prevention (including seat belts and bicycle and motorcycle helmets) has become a routine part of primary care practice.

TABLE 6-6

Preventive Counseling Recommended by the U.S. Preventive Services Task Force (USPSTF)

TOPIC	CHAPTER REFERENCE
Alcohol and drug use	**453, 456, 457**
Genetic counseling for *BRCA1/2* testing among women at increased risk for deleterious mutations	**79, 467**
Nutrition and diet	**332, 333**
Sexually transmitted infections	**136, 202**
Sun exposure	**61**
Tobacco use	**454**

IMPLEMENTING DISEASE PREVENTION AND SCREENING

The implementation of disease prevention and screening strategies in practice is challenging. A number of techniques can assist physicians with the delivery of these services. An appropriately configured electronic health record can provide reminder systems that make it easier for physicians to track and meet guidelines. Some systems give patients secure access to their medical records, providing an additional means to enhance adherence to routine screening. Systems that provide nurses and other staff with standing orders are effective for immunizations. The USPSTF has developed flow sheets and electronic tools to assist clinicians (*https://www.uspreventiveservicestaskforce.org/uspstf/information-health-professionals*). Many of these tools use age categories to help guide implementation. Age-specific recommendations for screening and counseling are summarized in **Table 6-7**.

TABLE 6-7

Age-Specific Causes of Mortality and Corresponding Preventive Options

15–24	1. Accident	• Counseling on routine seat belt use, bicycle/motorcycle/ATV helmets (1)

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

AGE GROUP	LEADING CAUSES OF AGE-SPECIFIC MORTALITY	SCREENING PREVENTION INTERVENTIONS TO CONSIDER FOR EACH SPECIFIC POPULATION
15–24	1. Accident 2. Homicide 3. Suicide 4. Malignancy 5. Heart disease	• Counseling on routine seat belt use, bicycle/motorcycle/ATV helmets (1) • Counseling on diet and exercise (5) • Discuss dangers of alcohol use while driving, swimming, boating (1) • Assess and update vaccination status (tetanus, diphtheria, hepatitis B, MMR, rubella, varicella, meningitis, HPV) • Ask about gun use and/or gun possession (2,3) • Assess for substance abuse history including alcohol (2,3) • Screen for domestic violence (2,3) • Screen for depression and/or suicidal/homicidal ideation (2,3) • Pap smear for cervical cancer screening after age 21 (4) • Discuss skin, breast awareness, and testicular self-examinations (4) • Recommend UV light avoidance and regular sunscreen use (4) • Measurement of blood pressure, height, weight, and body mass index (5) • Discuss health risks of tobacco use, consider emphasis on cosmetic and economic issues to improve quit rates for younger smokers (4,5) • Chlamydia and gonorrhea screening and contraceptive counseling for sexually active females, discuss STD prevention • Hepatitis B, and syphilis testing if there is high-risk sexual behavior(s) or any prior history of sexually transmitted disease • Hepatitis C screening starting at age 18 to 79 HIV testing • Continue annual influenza vaccination
25–44	1. Accident 2. Malignancy 3. Heart disease 4. Suicide 5. Homicide 6. HIV	*As above plus consider the following:* • Readdress smoking status, encourage cessation at every visit (2,3) • Obtain detailed family history of malignancies and begin early screening/prevention program if patient is at significant increased risk (2) • Assess all cardiac risk factors (including screening for diabetes and hyperlipidemia) and consider primary prevention with aspirin for patients at >3% 5-year risk of a vascular event (3) and statin therapy for higher risk patients • Assess for chronic alcohol abuse, risk factors for viral hepatitis, or other risks for development of chronic liver disease • Consider individualized breast cancer screening with mammography at age 40 (2)
45–64	1. Malignancy 2. Heart disease 3. Accident 4. Diabetes mellitus 5. Cerebrovascular disease 6. Chronic lower respiratory disease 7. Chronic liver disease and cirrhosis	• Consider prostate cancer screen with annual PSA and digital rectal examination at age 50 (or possibly earlier in African Americans or patients with family history) (1) • Begin colorectal cancer screening at age 45 or 50 with fecal occult blood testing, flexible sigmoidoscopy, or colonoscopy (1) • Reassess and update vaccination status at age 50 and vaccinate all smokers against *Streptococcus pneumoniae* at age 50 (6) • Consider screening for coronary disease in higher-risk patients (2,5) • Zoster vaccination at age 60 • Begin mammography screening by age 50 • Lung cancer screening at age 50 to 80 years if a 20 pack-year smoking history and currently smoke or have
	2. Malignancy 3. Cerebrovascular	• Readdress smoking status, encourage cessation at every visit (1,2,3,4) • One-time ultrasound for AAA in men 65–75 who have ever smoked

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

	disease and cirrhosis 8. Suicide	*pneumoniae* at age 50 (6) • Lung cancer screening at age 50 to 80 years if a 20 pack-year smoking history and currently smoke or have quit within the past 15 years, yearly.
≥65	1. Heart disease 2. Malignancy 3. Cerebrovascular disease 4. Chronic lower respiratory disease 5. Alzheimer's disease 6. Influenza and pneumonia 7. Diabetes mellitus 8. Kidney disease 9. Accidents 10. Septicemia	*As above plus consider the following:* • Readdress smoking status, encourage cessation at every visit (1,2,3,4) • One-time ultrasound for AAA in men 65–75 who have ever smoked • Consider pulmonary function testing for all long-term smokers to assess for development of chronic obstructive pulmonary disease (4,6) • Screen all postmenopausal women (and all men with risk factors) for osteoporosis • Continue annual influenza vaccination and vaccinate against *S. pneumoniae* at age 65 (4,6) • Screen for visual and hearing problems, home safety issues, and elder abuse (9) • Consider fall prevention exercise intervention if at higher risk (9)

Note: The numbers in parentheses refer to areas of risk in the mortality column affected by the specified intervention.

Abbreviations: AAA, abdominal aortic aneurysm; ATV, all-terrain vehicle; HPV, human papillomavirus; MMR, measles-mumps-rubella; PSA, prostate-specific antigen; STD, sexually transmitted disease; UV, ultraviolet.

Many patients see a physician for ongoing care of chronic illnesses, and this visit provides an opportunity to include a "measure of prevention" for other health problems. For example, a patient seen for management of hypertension or diabetes can have breast cancer screening incorporated into one visit and a discussion about colon cancer screening at the next visit. Other patients may respond more favorably to a clearly defined visit that addresses all relevant screening and prevention interventions. Because of age or comorbidities, it may be appropriate with some patients to abandon certain screening and prevention activities, although there are fewer data about when to "sunset" these services. For many screening tests, the benefit of screening does not accrue until 5–10 years of follow-up, and there are generally few data to support continuing screening for most diseases past age 75. In addition, for patients with advanced diseases and limited life expectancy, there is considerable benefit from shifting the focus from screening procedures to the conditions and interventions more likely to affect quality and length of life.

FURTHER READING

Bretthauer M et al: America, we are confused: The updated U.S. Preventive Services Task Force recommendation on colorectal cancer screening. Ann Intern Med 166:139, 2017. [PubMed: 27820949]

Hayes JH et al: Observation versus initial treatment for men with localized, low-risk prostate cancer: A cost-effectiveness analysis. Ann Intern Med 158:853, 2013. [PubMed: 23778902]

Hugosson J et al: Mortality results from the Goteborg randomized population-based prostate-cancer screening trial. Lancet Oncol 11:725, 2010. [PubMed: 20598634]

Oeffinger KC et al: Breast cancer screening for women at average risk 2015. Guideline update from the American Cancer Society. JAMA 314:1599, 2015. [PubMed: 26501536]

US Preventive Services Task Force: Screening for colorectal cancer. US Preventive Services Task Force recommendation statement. JAMA 315:2564, 2016. [PubMed: 27304597]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 6: Screening and Prevention of Disease, Katrina A. Armstrong; Gary J. Martin
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 7: Global Diversity of Health System Financing and Delivery

Richard B. Saltman

INTRODUCTION

Health care systems are highly complex organizations, with many interdependent components. In developed countries, health systems have traditionally been classified by their type of financing—i.e., either predominantly tax-funded (such as the National Health Service in England and publicly operated regional care systems in the four European Nordic countries) or predominantly statutory social health insurance (SHI)-funded (such as in Germany, the Netherlands, and France). Over the past several decades, however, there has been structural convergence in the technical characteristics of both funding arrangements and in the associated delivery systems, making analytic observations about differences across national systems more difficult.

A second confounding factor has been that former Soviet Bloc countries in Central and Eastern Europe, including the Russian Federation, have, since 1991, replaced their former Soviet-style Semashko models (a top-down, national government–controlled funding and delivery structure with a parallel Communist Party administrative apparatus) with various hybrid arrangements built on national government–run SHI financing. Distinctions across developed country health systems, especially in Europe, have been further compressed by inadequate resources in many publicly funded systems in an era of rapid clinical and technological change, triggering increased private sector funding and provision.

In middle-income developing countries, institutional structures in the health sector typically reflect the country's preindependence administrative framework. Mexico, for example, has a Spanish-derived configuration with health insurance as part of social insurance for formally employed workers (via Instituto Mexicano del Seguro Social), supplemented by tax-funded health services (Seguro Popular) provided for those with informal employment and all other citizens, as well as a separate program (Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado) for public employees. Countries such as India and Egypt, reflecting British influence, have predominantly tax-funded and publicly operated health systems. China is an exception, with an internally generated system that is publicly funded and operated, although recent Communist Party policy has been to introduce SHI-based insurance with individual medical savings accounts (patterned after Singapore), promote private insurance, and expand private hospitals.

In lower-income developing countries, health services are typically provided by tax-funded public institutions, often with considerable inadequacies and sometimes with substantial copayments. It is important to note that governmentally organized systems in nearly all developing countries, as well as in former Soviet bloc countries and, to a lesser degree, in tax-funded developed countries, are supplemented to varying extents by a mix of private and/or employer-paid insurers and providers.

This chapter focuses on the individual patient care system: on the financing and delivery of individual clinical and preventive services. The individual patient care system is composed of the financing and delivery of necessary services to prevent death or serious harm ("rule of rescue"); to maintain quality of life; and to manage, reduce, and/or prevent the burden of illness on individual patients. While the technical dimensions of most clinical services are similar across countries, their organizational, social, and economic characteristics range widely. Health systems in both developed and developing countries exhibit substantial differences, for example, in access to care; in the design and reliance on quality assurance and provider payment mechanisms; in the relationship of primary care to hospital services; in the coordination of health care with home care and nursing home services; in the design and use of provider management strategies; in the way physicians work and are paid; in the decision-making roles of politically elected officials and of national, regional, and municipal governments; and in participation of both citizens and patients. These wide-ranging institutional and organizational characteristics reflect differing country contexts (geographical, social, economic, and political), differences in national culture (consisting of prioritized norms and values), and substantial variation in how health sector institutions are structured.

FINANCING INDIVIDUAL PATIENT CARE SERVICES IN DEVELOPED COUNTRIES

Funding for individual care services in developed countries comes from the particular national mix among four possible sources of revenue: national,

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Total Health Expenditures

Funding for individual care services in developed countries comes from the particular national mix among four possible sources of revenue: national, regional, and/or municipal taxes; mandatory SHI; private health insurance (including employer-paid insurances); and out-of-pocket payments. Most countries have one preponderant payer, which then defines its funding arrangements and serves to frame the structure of its delivery system as well.

Total Health Expenditures

The Organization for Economic Co-operation and Development (OECD) data from 2017 (adjusted for purchasing power parities) show that total health care expenditures in developed countries vary across a considerable range, tied to health system structure as well as national history and culture (Table 7-1).

TABLE 7-1

Developed Country Total Health Expenditure (% GDP)

TAX FUNDED IN WESTERN EUROPE		SHI FUNDED IN WESTERN EUROPE		CENTRAL EUROPEAN		DEVELOPED ASIAN		DEVELOPED NORTH AMERICAN	
Ireland	7.2%	Belgium	10.3%	Latvia	6.0%	Singapore	4.5%	Canada	10.7%
Spain	8.9%	Netherlands	10.1%	Poland	6.5%	South Korea	7.6%	United States	17.1%
UK	9.6%	Germany	11.2%	Czech Republic	7.2%	Japan	10.9%		
Finland	9.6%	Switzerland	12.3%	Slovenia	8.2%				
Denmark	10.1%								
Sweden	11.0%								

Abbreviations: GDP, gross domestic product; SHI, social health insurance; UK, United Kingdom.

Source: The Organization for Economic Co-operation and Development (OECD) data.

Per capita health expenditure figures provide a different, specific measurement of available funds in a country's health sector (Table 7-2).

TABLE 7-2

Developed Country Per Capita Health Expenditures

TAX FUNDED IN WESTERN EUROPE		SHI FUNDED IN WESTERN EUROPE		CENTRAL EUROPEAN		DEVELOPED ASIAN		DEVELOPED NORTH AMERICAN	
Spain	$2738	Belgium	$4149	Latvia	$874	South Korea	$2043	Canada	$4458
Italy	$2738	Germany	$4714	Poland	$809	Singapore	$4083	United States	$9869
UK	$3958	Netherlands	$4742	Czech Republic	$1321	Japan	$4233		
Denmark	$5565	Switzerland	$9835	Slovenia	$1834				

Abbreviations: SHI, social health insurance; UK, United Kingdom.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 7-2

Developed Country Per Capita Health Expenditures

TAX FUNDED IN WESTERN EUROPE		SHI FUNDED IN WESTERN EUROPE		CENTRAL EUROPEAN		DEVELOPED ASIAN		DEVELOPED NORTH AMERICAN	
Spain	$2738	Belgium	$4149	Latvia	$874	South Korea	$2043	Canada	$4458
Italy	$2738	Germany	$4714	Poland	$809	Singapore	$4083	United States	$9869
UK	$3958	Netherlands	$4742	Czech Republic	$1321	Japan	$4233		
Denmark	$5565	Switzerland	$9835	Slovenia	$1834				
Sweden	$5710								

Abbreviations: SHI, social health insurance; UK, United Kingdom.

Tax-Funded Systems

In the United Kingdom, 79% of all health care funding was furnished through general tax revenues allocated by the national government in its annual budget process (all figures from OECD for 2017). In Sweden, all public taxes combined raised 83.7% of total health care spending. Sweden's 21 regional-level elected governments provide approximately 70% of that 83.7%, with the remaining 13.7% of total health spending raised by national and municipal taxes. In Canada, 71% of total health spending was raised by tax revenues, with 66% of that 71% coming from provincial or territorial taxes, while 5% came from national and local government taxes.

In most tax-funded countries, a segment of the population also has individual-, company-, or union-purchased private complementary and/or supplemental insurance coverage. In Sweden, 2019 estimates are that about 600,000 individuals have private complementary policies in a total population of 9 million. In Denmark, 50% of the population purchase supplemental insurance, while 30% have complementary insurance (often purchased by employers) that pays for private sector services enabling them to bypass public sector queues. In Finland, many middle-class families purchase separate private health insurance for their children to enable them to bypass long waiting times for primary and secondary pediatric health care services. More than 400,000 Finnish children (in a total population of 5 million) have privately purchased policies. In England in 2015, individual-, employer-, and union-purchased private complementary insurance covered an estimated 10.5% of the population, or about 6 million people. In Canada, individuals are not allowed by law to purchase private complementary insurance (except for Supreme Court–ordered insurance for three backlogged surgical procedures in Quebec Province—2005 Chaoulli decision); however, approximately 65% of the population have employer-, union-, or private group–purchased supplemental insurance for non–publicly covered services such as outpatient pharmaceutical prescriptions and home care.

Social Insurance–Funded Systems

In Western Europe, SHI funds have traditionally been organized on a private not-for-profit basis, but with statutory responsibilities under national law. When former Soviet Bloc countries in Eastern Europe regained their independence in 1991, they returned to pre–World War II SHI models, but because there was no remaining organizational infrastructure, these post-1991 arrangements typically became a single SHI fund, run as an arm of the national government. In the United States, the Medicare social insurance system for citizens over age 65, enacted in 1965, is organized as a single fund tied to the national Social Security (public pension) Administration, an independent agency within the national government, with reimbursement arrangements supervised by the Centers for Medicare and Medicaid Services (CMS) inside the Department of Health and Human Services. Medicare covers inpatient hospital care plus limited post-hospital nursing home services (Medicare Part A). Supplemental private insurance policies are bought by covered individuals to help pay for outpatient physician visits (Medicare Part B) and for outpatient pharmaceuticals (Medicare Part D)

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

enroll or remain. Eleven percent of the population—all having annual incomes above the mandatory SHI enrollment ceiling of 54,500 euros—have opted out of the SHI system to voluntarily enroll in claims-based private health insurance, whereas 4% of the citizenry is enrolled in sector-specific

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

there was no remaining organizational infrastructure, these post-1991 arrangements typically became a single SHI fund, run as an arm of the national

covers inpatient hospital care plus limited post-hospital nursing home services (Medicare Part A). Supplemental private insurance policies are bought by covered individuals to help pay for outpatient physician visits (Medicare Part B) and for outpatient pharmaceuticals (Medicare Part D).

In Germany, 85% of the population is enrolled in one of 120 not-for-profit, monthly premium–based private SHI funds. This figure includes all individuals with annual incomes below 54,500 euros, who are required by law to join an SHI fund, as well as those with higher incomes who choose to enroll or remain. Eleven percent of the population—all having annual incomes above the mandatory SHI enrollment ceiling of 54,500 euros—have opted out of the SHI system to voluntarily enroll in claims-based private health insurance, whereas 4% of the citizenry is enrolled in sector-specific public programs such as the military. Since 2009, all SHI members pay a flat tax on gross monthly income as a contribution (8.2% in 2018, up to an upper income limit of 49,500 euros), which is transferred by their SHI fund to a national pool, and then redistributed back to their chosen fund on an individual risk-adjusted basis. Employers send 7.3% of each employee's salary to the same national pool. Special arrangements exist for payments from self-employed, retired, and unemployed workers. Since 1995, there has been a separate mandatory social insurance fund for long-term care (LTC), with an annual premium of 1.95% of each adult's gross monthly income, split 50%–50% with their employer. Pensioners since 2004 are required to pay the full 1.95% from their pensions. Childless SHI enrollees pay a surcharge of 0.25% of monthly gross income. Overall, 78% of all health care expenditures in Germany were paid from public and/or mandatory private SHI sources.

In the Netherlands since 2006, all adult citizens pay a fixed premium (about 1453 euros in 2019) to their choice among 35 private health insurers (not-for-profit and for-profit), with four large insurance groups having over 1 million members each. In addition, employers pay 6.95% of salary below 51,400 euros for each employee into a national health insurance fund. Self-employed individuals pay 4.85% into the national fund for taxable income up to the same limit. Retired and unemployed individuals also make payments. In addition to the individual premiums paid to their choice of private insurance fund, payments from the national health insurance fund, adjusted by individual age, sex, and health characteristics, also are made to the individual's chosen insurer. The Netherlands has a separate mandatory social insurance fund for LTC (the ABWZ, since 2015 the WLZ, and now only for residential nursing home care) to which each employee pays 9.5% of taxable income beneath 33,600 euros every year. Self-employed, unemployed, and retired individuals also are required to pay premiums to the WLZ. Overall, including SHI revenues, public spending provided 87% of total health expenditures in 2014.

In Estonia, a former Soviet Republic that re-established an SHI system in 1991 upon regaining its independence, there is one national SHI fund that is an arm of the national government. This fund collects mandatory payments of 13% from salaried workers and 20% from self-employed individuals, covering both health care and retirement pensions. Overall, including SHI revenues, public spending accounted for 74.5% of total health expenditures in 2017.

Singapore, Japan, South Korea, and Taiwan have predominantly SHI systems of funding for individual care services. In these Asian countries (except Japan), there is one SHI fund that typically is operated as an arm of the national government.

In Singapore, starting in 1983, all employees up to age 50 have been required to place 20% of their income (employers add 16% more) into a personal health savings account to pay for direct health care costs, managed in their name by the Singapore government, called a Medisave account. Medisave accounts have a maximum amount, are tax-exempt, and receive interest payments (currently set at 4%). Consistent with a Confucian emphasis on family, the funds that accumulate in the Medisave account can be spent on health care for family members as well. If the accumulated funds are not spent on health care during the insured's life, they become part of the individual's personal estate and are distributed as a tax-free inheritance to his or her designated heirs. In addition, Singaporean citizens are also automatically enrolled into a second government-run health insurance plan called MediShield that pays for supplemental catastrophic, chronic, and long-term care. While citizens can opt out, 90% of citizens remain in the program. The Singapore government also operates a third, wholly tax-funded payer called Medifund that, with approval of a local neighborhood committee, will pay hospital costs for 3–4% of the population who are recognized as indigent. In part reflecting the high level of mandatory individual saving, public funding provided only 54.5% of total health expenditures in 2016.

In South Korea, a state-run SHI system was established in 1977, which in 1990 covered 30.9% of total health care costs. This percentage paid by the SHI system rose to 40.5% of total costs in 2017, with national tax revenue covering 16.9%, leaving out-of-pocket expenses at a relatively high 34.4% of total costs. Although there are legal ceilings on total out-of-pocket copayments for each 6-month period, over 70% of Korean adults purchase an additional private Voluntary Health Insurance policy to cover these additional direct expenditures. In 2000, three types of public SHI funds were merged into a single national state-run fund. As of 2018, 6% of an employee's salary must be paid as a social insurance contribution into this fund, with employees and employers each paying 50% of that amount. In 2008, an additional SHI fund was introduced to pay for LTC, operated by the main state-run SHI fund to reduce administrative costs. Contributions to the LTC fund are set at 6.55% of the individual's regular SHI contribution, coupled with 20% copayments for institutional care and 15% copayments for home care services.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

health insurance covered 34% of total health expenditures, Medicare (mandatory SHI program for all citizens over 65) covered 20%, Medicaid (a joint federal-state welfare program for low-income citizens) covered 17%, and out-of-pocket paid 10%. Sources of funds for these programs were 28% from

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

The United States

There is no single preponderant source of health care spending in the United States. The federal government's CMS reported that, for 2017, private health insurance covered 34% of total health expenditures, Medicare (mandatory SHI program for all citizens over 65) covered 20%, Medicaid (a joint federal-state welfare program for low-income citizens) covered 17%, and out-of-pocket paid 10%. Sources of funds for these programs were 28% from the federal government, 17% from state and local governments, 28% from private households, and 20% from private business (e.g., employers). The World Bank set public funding in the United States at 50.2% of total health expenditures in 2017.

In 2010, the passage of the Affordable Care Act (ACA) extended privately provided but heavily regulated and federally subsidized health insurance to many low- and middle-income uninsured individuals and families. Since the same act reduced the availability of existing individually purchased private health insurance, the total increase in the number of newly covered individuals was less than expected. Insurance premium increases for 2017 rose from 20% to over 100%, depending on the particular state, with additional increases in up-front deductible requirements, raising questions about the long-term sustainability of the ACA initiative. The recent Republican administration sought to repeal major financial and tax elements of the ACA and to replace existing subsidy arrangements with a system of refundable tax credits toward the establishment of individual health savings accounts and/or purchase of private health insurance on open cross-state markets (currently, private health insurance in the United States remains controlled at the separate 50-state level of government).

DELIVERING INDIVIDUAL PATIENT CARE SERVICES IN DEVELOPED COUNTRIES

Hospital Services

In Europe, hospitals in both tax-funded and SHI-funded health systems are mostly publicly owned and operated by regional or municipal governments. In tax-funded health systems, most hospital-based physicians are civil servants, employed on a negotiated salary basis (often by a physician labor union), and subject to most of the usual advantages and disadvantages of being a public sector employee. There are somewhat more private hospitals in SHI-funded health systems. However, most larger hospitals are public institutions operated by local governments, and most hospital physicians (with the notable exception of the Netherlands, where they are private contractors organized in private group practices) are, like those in tax-funded systems, public sector employees. In most tax-funded European countries (but not continental SHI-funded countries), few specialist physicians have office-based practices, and in both tax- and SHI-funded systems, office-based specialists do not have admitting privileges to publicly operated hospitals.

Most public hospitals in both tax-funded and SHI-funded health systems are single free-standing institutions that can be classified into three broad categories by complexity of patients admitted and number of specialties available: (1) district hospitals (four specialties: internal medicine, general surgery, obstetrics, and psychiatry); (2) regional hospitals (20 specialties); and (3) university hospitals (>40 specialties). In addition, many countries have a number of small, 15- to 20-bed, freestanding, private (typically for-profit) clinics. Recently, some tax-funded countries have begun to merge district and regional hospitals in an effort to improve the quality of care and create financial efficiencies (for example, Norway; planned for Denmark, also for Ireland; however, failed Parliamentary passage and brought down the coalition government, in Finland in 2019). Institutional mergers can be difficult to negotiate among publicly operated hospitals, due to the role that these large institutions play as important care providers and as large employers in smaller cities and towns, especially given political and union concerns about maintaining current employment levels. In the United States, financial and reimbursement pressures triggered by the implementation of the 2010 ACA have generated a number of private sector hospital mergers into larger hospital groups.

In tax-funded health systems, publicly funded patients who are admitted for an elective procedure cannot choose their specialist physician (except private-pay patients in "pay beds" in National Health Service [NHS] hospitals in England). Specialists are assigned by the clinic to a patient based on availability, with both junior and senior doctors placed in rotation.

Capital costs (buildings, large medical equipment) are publicly funded in all tax-funded systems and in most traditional SHI systems. For example, in Germany, capital costs for public hospitals are paid for by the regional governments. As a result, new capital investment is often allocated politically, according to location and political priorities. In Finland, local politicians in the 1980s would say that it "takes 10 years to build a hospital," meaning that it took that long to become a political priority for the regional government that controlled capital expenditures. Local politicians would therefore regularly overbuild when they got their one opportunity to obtain new capital.

Recently, efforts have been made to make public hospitals more responsible for their use of capital. In the Netherlands, public hospitals were shifted

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

build turn-key facilities (thus taking capital costs off the public borrowing limit), and then rent these facilities back to the NHS and/or the relevant NHS Foundation Trust. In Sweden and Finland, while capital equipment is now a cost on hospital operating budgets, large new capital equipment and

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Recently, efforts have been made to make public hospitals more responsible for their use of capital. In the Netherlands, public hospitals were shifted into private not-for-profit entities that are expected either to fund new capital from operating surplus or to borrow the funds from a bank based on a viable business plan. In England, more than 100 hospitals have been built using the Public Finance Initiative (PFI) program, in which private developers build turn-key facilities (thus taking capital costs off the public borrowing limit), and then rent these facilities back to the NHS and/or the relevant NHS Foundation Trust. In Sweden and Finland, while capital equipment is now a cost on hospital operating budgets, large new capital equipment and major building renovations remain politically driven processes often with extensive delays. In Stockholm County, the New Karolinska University Hospital opened in 2018 was built and is managed by a separate nonprofit public-private company.

In Singapore and South Korea, both of which are SHI funded, larger hospitals are publicly operated. However, there are a substantial number of smaller private clinics typically owned by specialist physicians. In the United States, the passage of the 2010 ACA has triggered the selling of many private specialist group practices to hospital groups, transforming previously independent practicing physicians into hospital employees.

Primary Care Services

Most primary health care in SHI-funded health systems, and also in an increasing number of tax-funded health systems (except in low-income areas of some large cities), is delivered by independent private general practitioners (GPs), working either individually or in small privately owned group practices. Recent changes in tax-funded health systems include Norway, where most primary care moved from municipally employed physicians to private-practice GPs in 2003, and Sweden, where, following a 2010 change in national reimbursement requirements, new privately owned not-for-profit and for-profit GP practices were established and now deliver 50% of all primary care visits.

In England, most primary care physicians are private GPs who are contractors to the NHS, working either independently or in small group practices. These private GPs own their own practices, which they can sell when they retire. However, as part of the original agreement to convince physicians to support the establishment of the NHS in 1948 (which most physicians strongly opposed), private GPs also receive a national government pension upon retirement. In the inner cities in England, there are some larger primary health clinics.

In 2001, England's private primary care doctors were organized into geographically based Primary Care Trusts (PCTs). These PCTs were allocated 80% of the total NHS budget to contract for elective hospital services required by their patients with both NHS hospital trusts as well as private hospitals. In 2013, PCTs were restructured into Clinical Commissioning Groups with similar contracting responsibilities.

In 2004, the Quality Outcomes Framework (QOF) was introduced as a quality of care–tied approach to providing additional income for NHS GPs. This regulatory mechanism in 2010 set 134 different standards for best practice primary care in four main domains: 86 clinical, 36 organizational, 4 preventive service, and 3 patient experience. GP income grew on average by 25% through the introduction of the QOF, with general practices averaging 96% of possible QOF points. Total spending on QOF in 2014 in England consumed 15% of all primary care expenditures.

In April 2019, a slightly revised QOF contract was implemented, which retired 28 low-value indicators, introduced 15 new more clinically appropriate indicators, added two Quality Improvement modules, and added a new personalized care adjustment option. Funding was only changed marginally.

Access for individuals to primary care services is considered good in SHI-funded systems such as those in Germany and the Netherlands. One often-cited reason is that private office-based physicians (both GPs and specialists) in these countries are paid on a modified fee-for-service basis. In Germany, office-based physicians are paid on a quarterly basis by the Sickness funds, acting jointly at the Länder (regional) level through a point-based system. A national agreement between the physician association and the association of sick funds establishes points for each clinical act. Similarly, the association of sick funds (led in each of Germany's 16 Länder by the fund with the most subscribers in that region) establishes a fixed budget for all office-based physician services for all sick fund patients each 3-month period. Retrospectively at the end of each period, the total number of points is divided into the sick funds' fixed allocation for office-based physicians for that Länder for that quarter, establishing the value of a point for that quarter. Subsequently, each office-based physician's point total is multiplied by that quarterly point value, resulting in that physician's total payment from the statutory sick funds.

In contrast to SHI systems, seeing a primary care doctor in a number of tax-funded health systems has become increasingly difficult over the past decade. In Sweden, in 2005, a "care guarantee" was introduced that required its predominantly publicly operated health centers to see a patient within 7 days after calling for an appointment. In Finland, where public primary health care centers used to provide most primary care visits, delays in getting public health center appointments have pushed up to 40% of all visits into a parallel occupational health system, as well as to publicly employed primary care physicians working privately in the afternoons.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

shortage of GPs, requiring large numbers of new nurse practitioners and other auxiliary personnel. In Central European countries that were formerly within the Soviet Bloc, primary care provision had to be newly established after independence was regained in 1991, since first-line care in the former

primary care physicians working privately in the afternoons.

In England in 2019, access to GP services has been labeled a "crisis," aggravated by a 6% fall in the number of practicing GPs, leading to delays of up to 30 days for an appointment in urban areas like London. A 2019 report by the King's Fund found that only 1 in 20 trainee GPs planned to work full time. Also in 2019, the Nuffield Trust published a report suggesting that future planning for primary care services in England should assume a permanent shortage of GPs, requiring large numbers of new nurse practitioners and other auxiliary personnel. In Central European countries that were formerly within the Soviet Bloc, primary care provision had to be newly established after independence was regained in 1991, since first-line care in the former Semashko model was provided in specialist polyclinics. Primary care doctors rapidly emerged as almost entirely private for-profit GPs, working on contract from the national SHI fund (Estonia, Hungary, North Macedonia), from state-regulated private insurance companies (Czech Republic), or from regional/municipal public payers (Poland). Private GPs in most Central European countries now are paid on a per-visit-tied basis. This arrangement was heavily influenced by the structure of primary care in Germany, where private office-based GPs are paid according to a point-system-tied framework.

In Asian countries such as Singapore, South Korea, and Japan, most primary care is provided by private for-profit GPs working independently or in small group practices. Private GPs are reimbursed at a set per-service fee by the national SHI fund(s). Access to primary care physicians is considered good.

Developed countries have varying policies regarding access to individual preventive services. Health systems in most countries provide vaccinations and mammography as part of funded health care services. In the United States, most insured individuals—and in Canada, most covered residents—automatically receive an annual physical exam including full blood profiles. In Norway and Denmark, adult physical exams are provided only upon special request by the individual, and in Sweden, adult physical exams are provided only to pregnant women. In Sweden, adults who wish to know their cholesterol or prostate-specific antigen (PSA) levels have begun to purchase blood tests out-of-pocket from private laboratories. In England in 2019, the NHS announced it would stop providing PSA screening tests for prostate cancer, even to men who requested one, similarly forcing concerned patients to purchase private laboratory testing.

Patients must make copayments to see a primary care doctor in some tax-funded health systems and in most SHI countries. In tax-funded systems, for example, Swedish patients are required to make a county-council-set copayment for each primary care visit up to a national-government-set annual ceiling, after which ambulatory visits (both primary and outpatient specialist) are not charged. Finland has a fixed copay for public health center visits, while Denmark's private GP visits do not have a copayment. In England, there is no copayment for GP visits.

In SHI health care systems in Europe and in Asia, patients usually are responsible for a copayment for both primary and office-based specialist care. To defray these charges (and to pay for other nonfunded services), a high percentage of citizens typically purchase additional supplemental health insurance. In France, where 95% of patients in 2015 purchased private supplemental insurance, patients paid directly the full fee for 65% of outpatient primary and specialist services, reimbursed subsequently by both their SHI fund and their supplemental insurance carrier for all payments (after deductibles), while for 35% of services (for low-income individuals and certain high-cost procedures), full agreed prices were paid directly to providers by SHI.

Access to Elective Specialist Care

Approximately half of all European health care systems have a gatekeeping system that requires referrals from primary care physicians in order to book hospital specialist visits (for publicly paid visits). In most tax-funded health systems (although not in most SHI systems), there are substantial waiting times, typically several months or more, for elective specialist appointments as well as for high-tech diagnostic and treatment procedures. Waiting times can be particularly long for cancer and other elective surgical or high-demand services. In Sweden, government figures from the summer of 2017 showed that, nationally, only 5–10% of prostate cancer operations were performed within 60 days after diagnosis.

In the English NHS, waiting lists for elective surgery in 2019 were often 6 months or longer. In August 2017, there were over 4,000,000 patients on NHS waiting lists. In January 2018, what administrators termed "a severe flu season," during which hospital emergency rooms were overwhelmed with elderly patients requiring admission, led to a national-level NHS decision to cancel all elective operating room procedures in all hospitals in England (>50,000 procedures in 1200 hospitals) for the entire month of January, further lengthening waiting lists. Regarding quality of care, again in England, a March 2018 report from the national Office of Health Economics found that, in 2016 and 2017, up to three-quarters of patients who could have undergone keyhole procedures were forced to undergo open surgery, resulting in an estimated 1 million procedures each year that were more invasive than clinically necessary.

Delays in some tax-funded systems also are procedural. In England, for example, a patient who requires a further consultation with a second specialist

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

There is also substantial waiting time for radiologic imaging services in most tax-funded systems. In Malta, the tax-funded health system's recent efforts to prioritize elective MRI investigations have succeeded in reducing waiting times from 18 months to 4 months. In both the Alberta and British

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Delays in some tax-funded systems also are procedural. In England, for example, a patient who requires a further consultation with a second specialist typically has to return to their primary care physician for a second referral and then has to wait in the regular patient queue for that second appointment.

There is also substantial waiting time for radiologic imaging services in most tax-funded systems. In Malta, the tax-funded health system's recent efforts to prioritize elective MRI investigations have succeeded in reducing waiting times from 18 months to 4 months. In both the Alberta and British Columbia Provinces in Canada, waiting times for a publicly funded nonemergency MRI can extend up to several months, whereas privately paid MRIs were available in both provinces within 1 week.

This issue of waiting times for specialist services in tax-funded health systems reflects a combination of growing demand (increasing/aging populations and changing clinical indications), financial constraints, and insufficient capacity, including inadequate physician working hours. For example, in the 1980s, when several surgical procedures for the elderly became more routine practice (e.g., hip replacement, coronary artery bypass graft, corneal lens implantation), the waiting list problem worsened. It had been mitigated somewhat through increased service capacity by the early 2000s, only to return as a growing policy challenge once public sector financial resources became constrained again after the 2008 global financial crisis. Timely cancer diagnosis and care continue to be a particularly sensitive issue, with tax-funded systems often taking several months for a patient to see an oncologist and then months more to begin treatment. In 2013 in Sweden, a newspaper journalist set off a political storm when he described women patients in one large county council (Malmo) who had to wait more than 40 days to receive the results from their breast cancer biopsy. In September 2019 in England, only 76.9% of patients with suspected cancer began treatment within 2 months of an urgent referral from a GP.

In response to pressure from national patient associations, a number of tax-funded health care systems introduced maximum waiting times for elective hospital procedures in the early 2000s. (Most Western European SHI systems do not have long waiting times or treatment guarantees for hospital care.) These maximum waiting times typically include initial primary care visits as well as specialist evaluations and treatment. In Denmark, a patient has the right to go to a different Danish public hospital for care after waiting 30 days without treatment. In Sweden, under the 2005 "waiting time guarantee," an untreated patient's local county council is required to pay for care in another county's hospital after 180 days. In a parallel process at the European Union (EU) level, beginning in 1997, the EU Court of Justice steadily expanded the right of all EU citizens to travel to another EU country in order to receive "timely" care, with their home country health system required to pay for that care.

In private not-for-profit SHI-funded health systems such as in Germany and Switzerland, waiting times for specialist visits and hospital procedures are typically a few weeks to 1 month. In the SHI system in France, which is more centrally organized and funded (part of the Napoleonic tradition of public administration), ongoing disputes about insufficient central government funding for public hospitals and staff salaries led in March 2019 to 9 months of hospital staff strikes, particularly in accident and emergency departments. In November 2019, the national government announced that it would take over 10 billion euros in public hospital debt as part of an effort to reverse staff cutbacks, bed and operating theater closures, and personnel flight to the private sector.

Long-Term Care Services

LTC (consisting of residential and home-based services) consumes a relatively small but increasing proportion of gross domestic product (GDP) in developed countries. In 2016, Norway (2.95% GDP), Sweden (2.87% GDP), and the Netherlands (2.64% GDP) all spent more than one-fourth of their total health expenditures on LTC (Eurostat and OECD figures). More than one-fifth of all health care expenditures went to LTC in Belgium (2.16% GDP), Ireland (1.55% GDP), and Denmark (2.5% GDP). Lower-spending countries included the United Kingdom (18% of health expenditures; 1.75% GDP), Germany (12% of expenditures; 1.33% GDP), and Spain (9% of expenditures; 0.81% GDP). In the United States, official figures put total LTC expenditures in 2016 at 4.9% of total health expenditures, or 0.9% of total GDP. (Note that these figures do not include emergency, inpatient, or outpatient hospital costs generated by elderly patients.)

Since nursing home care is more expensive than home care (nursing home care requires the provision of housing, food, and around-the-clock care providers), government policymakers seek to keep the elderly and the chronically ill out of nursing homes for as long as feasible. Moreover, in developed countries like Sweden, Norway, and the United States, some 70% of all home care services come from informal caregivers: spouses, children (typically daughters), neighbors, and nonprofit community groups. While some SHI systems (e.g., Germany) have separate public LTC insurance (funded by mandatory premiums paid by all adults) that make available cash payments for LTC that can be used to compensate informal caregivers, most policymakers work hard to not monetize what is a large amount of essentially free care. Indeed, policymakers actively seek to encourage those providing these services to continue to do so as long as possible, trying to postpone caregiver burnout by providing support services such as free respite care, special call-in lines for caregiving advice, pension points toward retirement for the informal caregiver (Nordic countries), and

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

these services are also delivered mostly by municipal employees, working according to union-negotiated protocols. In some European SHI systems, and recently in tax-funded Sweden and Finland, private companies also provide home care services on contract to municipal governments. In

developed countries like Sweden, Norway, and the United States, some 70% of all home care services come from informal caregivers: spouses, encourage those providing these services to continue to do so as long as possible, trying to postpone caregiver burnout by providing support services such as free respite care, special call-in lines for caregiving advice, pension points toward retirement for the informal caregiver (Nordic countries), and free day-care center services.

In most tax-funded and SHI-funded European countries, home care services are organized at the municipal government level. In tax-funded systems, these services are also delivered mostly by municipal employees, working according to union-negotiated protocols. In some European SHI systems, and recently in tax-funded Sweden and Finland, private companies also provide home care services on contract to municipal governments. In combination with national legislation, these municipal systems also provide important support for informal caregivers, since the financial costs of caring for adults in their own home are substantially less than providing housing, food, and caregiver support in publicly funded homes for the aged or in nursing homes.

A high proportion of nursing homes in European tax-funded and SHI-funded health systems are publicly owned facilities operated by municipal governments; in some instances, in SHI-funded systems (Israel, the Netherlands), they are operated by private not-for-profit organizations. Recently, in some tax-funded systems (e.g., Sweden), private for-profit chains have begun to open nursing homes that are funded on a contract basis with local municipal governments. Costs for nursing home care can be expensive: in Norway, the cost per patient is often over $100,000 per year in a publicly funded home, with the patient responsible for paying up to 80% depending on the family's economic status. In Sweden, patients living in publicly funded nursing homes in Stockholm County pay a relatively small official fee, but they also pay room rent and up to 2706 Swedish krona (SEK) per month (about $270 U.S. dollars [USD]) for food out of their monthly public pension payments.

In 2012, in an effort to reduce demand for expensive hospital and nursing home services, Norway and Denmark began elderly care reforms that shifted service delivery as well as funding responsibilities to municipal governments. Among innovations in Norway, municipalities are required to establish a municipal acute bed unit (MAU) to treat stable elderly patients and provide observation beds for evaluation. Partial funding for these units is provided by the four public regional health care administrations. Some municipalities have also embedded primary care units inside their regional hospital to arrange discharge and to coordinate care for the chronically ill elderly. Norwegian municipalities are also responsible through their contracted (mostly private) primary care physicians to implement the National Pathways Program, which established treatment protocols for cross-sector conditions such as diabetes and cardiovascular conditions.

A differently configured structural innovation to better integrate LTC for the chronically ill elderly with clinical individual health services has been to consolidate both social and health care services within the same public administrative organization. In 2019, as part of health reforms in Ireland and Denmark and a proposed (unenacted) reform in Finland, as well as a pilot decentralization program in England for 2.8 million people in Greater Manchester, social and health care programs are to be administered by a single responsible agency.

In the SHI-funded system in the Netherlands, almost 7% of the population live in a residential home. National government legislation revised the structure of nursing home funding and care in 2015. Three acts restructured the separate public LTC SHI fund, which requires mandatory payments by 100% of Dutch adults, and introduced delivery-related reforms that reduced the number and overall cost of nursing home patients paid for by the fund. Determination of eligibility for public payment for nursing home care is now made by an independent national assessment body (the Centre for Needs Assessment). Moreover, municipal governments now play a stronger role in funding and delivering home care services. The reforms created social care teams that hold "kitchen table talks" to steer the elderly first toward seeking care from family, neighbors, churches, and other local community organizations before they qualify for publicly paid in-home care. In 2012, some 1.5 million people (12% of total population) provided informal care to ill or disabled persons, averaging 22 hours per week of care per person.

Home care recipients in the Netherlands can choose to set up a "personal budget," using their public funding allocation to select their preferred individual care personnel (either publicly employed or publicly approved private providers). This arrangement also enables these home care recipients to determine the particular mix of services they want, as well as to augment the allocated public funds with personal funds. A number of innovative not-for-profit nursing homes have been created to provide additional services to elderly living in their neighborhood (primary care home visits), as well as terminal hospice care (e.g., the Saffier De Residentie Groep residences in The Haag).

In the United States, nursing home and home care are funded and delivered in a variety of different ways. For individuals who have minimal financial assets, nursing home costs are paid by a joint federal-regional (state) welfare program called Medicaid. Most state government Medicaid programs pay out more than 40% of their total budget for nursing home care. In the past, Medicaid did not pay for home care services. However, some states have programs with private for-profit and not-for-profit providers that provide home care as a way to forestall the need for the more expensive nursing home care.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

affordable. The 2010 ACA contained a new public LTC insurance program. However, the program was designed to be voluntary, and U.S. Department of Health and Human Services administrators decided in 2013 not to implement that portion of the law.

Many private individuals take out private LTC insurance, typically from commercial insurance companies. These policies require individuals to make premium payments for years in advance (often 20 or more) before the individual learns whether they will, in fact, require home or nursing home care. Some private insurers have also raised premiums after individuals have paid in for many years and canceled policies if the new higher rate is not affordable. The 2010 ACA contained a new public LTC insurance program. However, the program was designed to be voluntary, and U.S. Department of Health and Human Services administrators decided in 2013 not to implement that portion of the law.

In addition to the tax-funded Medicaid program and privately purchased LTC insurance, many middle-class families pay for care from savings, by selling the elderly person's home, or by direct contribution from children and other family members. Expenses can reach between $60,000 and $100,000 per year depending on the location of a facility and who operates it.

Nursing home care in the United States is provided by a wide mix of private not-for-profit and for-profit providers, ranging from church-owned single-site homes to large stock market–listed companies. Many of these homes are purpose-built as assisted-living or memory-care facilities. Home care services are delivered by a mix of private not-for-profit and for-profit providers.

In Japan, a national LTC insurance fund was introduced in 2000. Although the new fund applies uniformly across the country, the program is administered by municipal governments and the premium level differs across municipalities, with an average monthly premium of 3000 yen (about $30 USD). In South Korea, an SHI fund for LTC is funded by mandatory contributions of 4.78% of a person's regular national health insurance contribution, with an additional 20% of total LTC expenditures provided by national government funds. The client copayment for home care is set at 15% of expenses and at 20% for residential care.

PHARMACEUTICALS

Pharmaceutical expenditures in developed countries (inpatient and outpatient combined) vary widely across different health system types, as well as between different countries within each institutional type. OECD figures for 2018 show drug expenditures in tax-funded countries in Western Europe ranging from 6.3% of total health expenditures (THE) in Denmark to 11.9% of THE in the United Kingdom and 18.6% of THE in Spain. In SHI-funded Western European systems, pharmaceuticals absorbed 7.5% of THE in the Netherlands, while in Germany, that figure was 14.1%. In the hybrid tax-funded SHI systems of Central Europe, the pharmaceutical percentage of THE is higher: 18.2% of THE in Estonia to 27.9% of THE in Hungary. Similarly, in Asian SHI systems, pharmaceuticals consumed 20.7% of THE in South Korea and 18.6% of THE in Japan. The OECD's 2018 figures for pharmaceutical spending in North America are 12.0% of THE in the United States and 16.7% in Canada.

Contributing factors to this wide-ranging variation are (1) differences in national practice and prescription patterns reflecting differing cultural expectations; (2) the ratio problem (relatively fixed level of pharmaceutical costs due to international prices—the numerator—divided by a greatly varying per capita health expenditure cost in different developed country health systems); (3) the range and type of pharmaceutical price controls in each country; and (4) the degree of limitation placed on pharmaceutical supply, tied to formularies and/or explicit forms of drug rationing.

Most European health systems have tight national controls on the cost and, in some tax-based countries, on the availability of pharmaceuticals. Most European countries also use a number of different regulatory measures to limit prices and/or availability of both inpatient and outpatient drugs, including mandatory generic prescribing, reference pricing, patient copays (sometimes with an annual ceiling, after which copayments are no longer required), and (particularly in tax-funded systems) national formularies tied to clinical effectiveness. Norway, for example, allows only about 2300 different preparations—including dosage, delivery method, and box size—to be stocked by pharmacies. Prices for drugs can vary considerably across different European countries, tied to economic development and domestic pricing patterns. One consequence of these differential national pricing controls has been the development of a parallel import market, in which drug wholesalers and pharmacists in the more expensive countries purchase supplies from a cheaper market elsewhere in Europe.

Access to expensive drugs has also been intentionally limited in some tax-funded health systems in Europe. One basis for rationing has been rationing tied to quality-adjusted life-years (QALYs). Rationing also reflects a clash between strained public drug budgets and public pressure. For example, in the case of cancer drugs in England, the recommendation of the National Institute for Health and Care Excellence (NICE) against funding the breast cancer drug trastuzumab (Herceptin) was subsequently overturned by the Minister of Health. Expensive cancer drugs continue to be rationed in England where the NHS Cancer Drug Fund, established in 2011 to provide access to non-NHS-provided drugs on a case-by-case basis, ran out of funds in 2015, forcing it to drop 25 of 83 covered drugs and close down for 3 months to restructure its operations.

As part of earlier medical patterns in Asian countries, office-based physicians traditionally filled prescriptions as well as prescribing drugs to patients.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

prescriptions dispensed in 2016 by physicians to 26% of prescriptions filled.

As part of earlier medical patterns in Asian countries, office-based physicians traditionally filled prescriptions as well as prescribing drugs to patients. These sales also served to supplement their income in a setting of relatively low per-visit payments from state-run SHI funds. Concerned about cost and overuse, both Taiwan (in 1997, except for emergency cases or rural regions) and South Korea (for the whole country in 2003) implemented "separation reforms," which ended these physician sales. In Japan, a series of fee and reimbursement reforms have trimmed the percentage of all prescriptions dispensed in 2016 by physicians to 26% of prescriptions filled.

GOVERNANCE AND REGULATION

Health care services in developed countries are steered, constrained, monitored, and (to varying degrees) assessed by governments and governmentally established and/or empowered bodies. Although these measures apply particularly to the financial efficiency of government-funded services, they also seek to promote patient and community safety, equity of access, and high-quality clinical outcomes. This oversight is often strongly focused on privately operated and contracted providers and insurers, although in principle, it applies to publicly operated organizations as well.

Governance consists of macro national-level policy, meso institutional-level management, and micro clinic-level care decisions. This complex mix of governance decisions is often shared among different national, regional, and local governments, depending on the degree of centralization, decentralization, or, recently, recentralization (e.g., Norway and Denmark). While most systems officially prioritize "good governance," governance activities frequently comingle with political objectives as core policy concepts are developed and transformed into concrete organizational targets.

In Sweden, health system governance is shared among national, regional (county), and local municipal governments. The national government has responsibility to pass "frame" legislation, which establishes the basic structure of the system. To cite one example, until recently, the national government had limited an adult patient's total copayments for outpatient physician care (specialist and primary care) and pharmaceuticals to 2800 SEK (about $280 USD) for a 12-month period. The 20 regional governments, in turn, made policy decisions within that legislation, deciding how to apportion the specific copayments for each primary care and specialist outpatient visit. Since Swedes can self-refer to specialists, some counties double the copayment to hospital-based doctors to discourage unnecessary appointments. Similarly, fiscal policy normally is shared between the regional government, which raises about 70% of total health expenditures through its own county-set flat income tax, and the national government, which provides additional purpose-tied funds for national objectives such as consolidating open-heart surgery across county lines as well as supplementing lower tax receipts in rural counties with smaller working populations. However, this normal funding relationship across governments can change. In the early 1990s, the national government placed a "stop" on raising county taxes prior to Sweden's admission in 1995 to the EU. In 2016, each of the 20 counties could set their own ceilings, which were almost all at 3300 SEK (about $330 USD).

In Spain's tax-funded health system (71.1% publicly funded in 2015), 17 regional "autonomous communities" were given full managerial responsibility for the provision of health services in a 1990s decentralization process, along with ownership of all publicly operated hospitals. The national government generates a substantial proportion of health care resources, which are included in the broad block grants it allocates to the regional governments, which then add regional tax revenue to make up the full public sector budget. In a mechanism to steer regional government operating policies in this decentralized environment, the national Spanish government established a joint federal-regional council to review quality and performance data (through the 2003 Health System Cohesion and Quality Act). Italy's tax-funded health system (75.8% publicly funded in 2014) similarly shares governance responsibilities between national and regional governments. Health services are provided by local health authorities (Azienda Sanitaria Locale) supervised by 20 regional governments within a nationally established governance framework, financed through a complicated mix of national and nationally stipulated but regionally collected taxes. Again, like Spain, the national government established a federal-regional government council, seeking to better coordinate care standards and information among the regions and with national government agencies. In 2006, the national government imposed strict financial plans on 10 regions that were systematically in deficit.

In Germany, where funding for its SHI-based health system is predominantly the responsibility of 120 private not-for-profit sickness funds, governance decisions are shared among these private sector sickness funds and public sector national, regional, and municipal governments. The sickness funds receive a risk-adjusted premium payment for each enrolled individual, according to a national government–determined formula, and from a national government–run health insurance pool. Most hospitals are owned and operated by municipal governments, while investment capital for structural renovations and new building comes from the 16 regional Länder taken from their tax revenues. Payment frameworks and amounts for public hospitals are negotiated between associations of these municipally owned hospitals and associations of the private sickness funds, without formal government participation.

Regulation is an essential element of an effective health care system and a key component of overall health system governance. Regulation

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

1. Requiring all cancer drugs adopted for use in the NHS to cost no more than $41,268/QALY;

Regulation is an essential element of an effective health care system and a key component of overall health system governance. Regulation incorporates both broad standard requirements that affect all organizations that operate in a country (e.g., hiring, firing, and wage decisions) as well as specific health sector–related regulations (e.g., proper handling, use, and disposal of low-grade nuclear waste from radiation treatments). Recent examples of health sector regulation in England, for example, include the following:

1. Requiring all cancer drugs adopted for use in the NHS to cost no more than $41,268/QALY;
2. Requiring in their employment contract that junior doctors in hospitals work a specific number of Sundays; and
3. Requiring that all emergency department patients receive care within 4 h of their arrival.

A powerful tool that has the force of law, regulation can have substantial negative as well as positive effects. A well-known political science corollary of regulatory power is that "the right to regulate is also the right to destroy." For example, in the United States, the federal Environmental Protection Agency, as part of its pursuit of cleaner air, issued wide-ranging regulatory orders setting performance standards that resulted in the closing of many West Virginia coal mines, with the loss of tens of millions of dollars of productive capacity and thousands of high-paying jobs, and likely contributing to social conditions that helped spawn that state's high rates of opioid abuse among unemployed males. Similarly, in some tax-funded European systems, such as those in Sweden and England, there is growing pressure from public health advocates for national regulations to prohibit the making of a profit from publicly paid funds. In Sweden, the national government's Reepalu report in 2016 honored a pledge made by the Social Democratic government to its Left (socialist) Party ally by calling for a legislated ban on profit-making in the provision of publicly funded health care services. The report's publication triggered substantial divestment of existing investor-owned primary care, nursing home, and home care companies.

FUTURE CHALLENGES

Health systems in developed countries face continued challenges in the coming years. These include financial, organizational, and policy dilemmas for which institutionally viable, financially sustainable, and politically supportable solutions will be complicated to develop and difficult to implement. On the delivery side, a key question is whether privately structured GP-based primary care is more efficient and effective than various clinic-based forms of primary care services. Recent movement in Northern and Central Europe toward more private GPs, along with continued private office–based primary care in much of Canada, the United States, and economically developed countries in Asia, raises complex policy issues for international organizations like the World Health Organization (WHO), as well as national policymakers. In the hospital sector, existing levels of clinical quality and patient responsiveness in publicly operated command-and-control institutions will increasingly have to compete with those of semi-autonomous public hospitals, as well as various types of private, sometimes very innovative providers. In the financing arena, continued pressure on publicly raised health system revenues is likely to erode longtime commitments in some tax-funded health systems to minimal patient copayments and low out-of-pocket funding.

An additional set of challenges will arise from recent commitments by international organizations like WHO to restructure health systems in developed countries to better address the social determinants of health. This new, incomplete strategy calls for a dramatic expansion of health sector responsibility to include a wide range of existing institutional arrangements in housing, education, work-life, and social and political decision-making. The influential 2010 Strategic Review of Health Inequalities in England entitled "Fair Society, Healthy Lives," led by Sir Michael Marmot, a British epidemiologist, called for the elimination of all "inequities in power, money, and resources." Separate from the political dimensions of this proposed new paradigm, how such fundamental societal change will be funded and implemented has yet to be addressed.

Looking forward, among the most essential challenges to national decision-makers in the coming period will be four specific health system imperatives:

1. **Finding a more sustainable balance between ethics and funding.**

   Policymakers in publicly funded health systems face a growing gap between patient expectations of high-quality clinical care, staff expectations of better compensation, and the economic imperative of no new taxes. Recent research has suggested that SHI-funded health systems, faced with increasing aging and thus proportionally fewer employed, face a similar gap. While the present solidaristic foundation for raising collective revenues is insufficient, available nonsolidaristic tools (copayments, supplemental insurance, private pay) inevitably contribute to overall inequality. But what then are the realistic policy alternatives? The minimalist new policy goal necessarily will have to become one of raising new revenues while doing the least economic and social harm.

2. **Developing better strategies to steer provider diversity**

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

   community groups, foundations, and cooperatives, as well as for-profit small local entrepreneurs, large international companies, and risk capital funds (venture capital). Furthermore, new innovative delivery models are reorganizing traditional service boundaries: not-for-profit private

2. **Developing better strategies to steer provider diversity.**

   Health systems in developed countries are becoming more diverse with more and different types of public owners: hospital trusts, state enterprises, and mixed public-private hospital owners/managers. There also are more and different types of private providers: not-for-profit community groups, foundations, and cooperatives, as well as for-profit small local entrepreneurs, large international companies, and risk capital funds (venture capital). Furthermore, new innovative delivery models are reorganizing traditional service boundaries: not-for-profit private nursing homes in the Netherlands also provide outpatient primary care to neighborhood elderly patients, as well as hospice care; Israeli technology companies combine high-tech home-based patient monitoring with standard medical and custodial home care services. Public pressure from citizens for more choice and better outcomes will pressure policymakers toward new, more accommodative health system arrangements. A 2019 national government report in Sweden on the hospital sector recommended a new emphasis on better access to out-of-office hours and out-of-hospital acute care by private as well as public providers.

3. **Ensuring better coordination between social and health services.**

   Tax-funded and SHI-funded systems alike are under intense policy pressure to develop better strategies to integrate services for the chronically ill elderly, as a way to improve the quality of services that these patients receive and to keep them at home healthier and longer, reducing expensive acute visits to hospitals and emergency departments. The clear delivery system goal will increasingly be to keep the elderly out of nursing homes and acute care facilities for as long as possible.

4. **Building labor unions into provider innovation.**

   In many developed countries, health sector staff, including hospital physicians, are members of labor unions. Effective policymaking will require finding mechanisms to build these personnel unions into accelerated health system restructuring processes. This process will necessarily involve integrating unions into more innovative, flexible, fiscally sustainable organizational arrangements with contracts that reward active participation in organizational change, contracts that pay incentives to more productive employees, quicker reassignment and redundancy procedures (firing health sector workers can take a year or longer in some European health systems), and establishing profit-sharing payments to teams/unions, also in public sector organizations.

While the structure and complexity of resolving these specific organizational challenges will vary depending on a country's cultural and institutional context, the commonality of these problems suggests that health systems in the developed world require a new, broader range of targeted policy strategies and solutions.

FINANCING AND PROVIDING HEALTH SERVICES IN DEVELOPING COUNTRIES

Health systems in developing countries reflect a complex combination of the same core elements found in developed country systems (hospitals, primary care facilities, medical staff, pharmaceuticals) adapted to different, widely varying organizational, social, political, and economic contexts and conditions **(See also Chap. 474)**. System structure and provider institutions typically vary by differing national characteristics including historical relationships (Anglophone/Francophone/Hispanic/Soviet Semashko/American institutional and educational links); GDP and per capita annual national income (low- or middle-income developing countries); political norms and values; and ethnic and/or cultural mix. Predominantly public sector funding, particularly in lower-income countries, typically generates substantially lower levels of resources per capita than in developed countries and tends to be less reliable, particularly in countries where the economy is dependent on commodity exports.

Service delivery arrangements in developing countries, in turn, typically have higher provider-to-population ratios as well as, in public sector institutions, more mixed quality of care. In a number of middle-income developing countries, migration of trained medical staff to practice in higher-paying developed country health systems (often going to countries with historical relationships and/or where they received advanced training) further depletes available medical resources. In nearly all developing countries, private sector providers play an important supplemental role, with some middle-income developing countries like China currently encouraging their further development.

Most middle- and lower-income developing countries struggle to fund high-quality individual health services. Recent emphasis on universal health coverage has intensified that struggle. In middle-income developing countries **(Table 7-3)**, World Bank data from 2016 show a range of health expenditure rates as a percentage of GDP, including Kazakhstan at 3.53% of GDP, Thailand at 3.71%, Malaysia at 3.80%, Turkey at 4.31%, China at 4.98%, Botswana at 5.46%, Mexico at 5.47%, and Colombia at 5.91%. Total health spending in low-income developing countries **(Table 7-4)** ranges

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

4.98%, Botswana at 5.46%, Mexico at 5.47%, and Colombia at 5.91%. Total health spending in low-income developing countries (**Table 7-4**) ranges from 3.65% of GDP for Nigeria, 3.66% for India, 3.97% for Ethiopia, 6.29% for Nepal, to 8.40% for Honduras.

TABLE 7-3

Middle-Income Developing Countries: Total Health Expenditure (% of gross domestic product)

Middle-Income Developing Countries	
Kazakhstan	3.53%
Thailand	3.71%
Malaysia	3.80%
Turkey	4.31%
China	4.98%
Botswana	5.46%
Mexico	5.47%
Colombia	5.91%

TABLE 7-4

Low-Income Developing Countries: Total Health Expenditure (% of gross domestic product)

Low-Income Developing Countries	
Nigeria	3.65%
India	3.66%
Ethiopia	3.97%
Nepal	6.29%
Honduras	8.40%

Given lower aggregate GDP levels, per capita annual expenditures are considerably less than those found in developed countries. In middle-income developing countries (**Table 7-5**), Thailand spent (2016 data in adjusted USD) $221 annually per person, Kazakhstan spent $262, Colombia spent $340, Malaysia spent $361, Botswana spent $379, China spent $398, Mexico spent $461, and Turkey spent $468. Among low-income developing countries (**Table 7-6**), Ethiopia spent $27 per person annually, Nepal spent $45, India spent $62, and Nigeria spent $79, whereas Honduras spent $199.

TABLE 7-5

Middle-Income Developing Countries: Per Capita Health Expenditures

Middle-Income Developing Countries	
Thailand	$221
Colombia	$340

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

countries (Table 7-6), Ethiopia spent $27 per person annually, Nepal spent $45, India spent $62, and Nigeria spent $79, whereas Honduras spent $199.

TABLE 7-5

Middle-Income Developing Countries: Per Capita Health Expenditures

Middle-Income Developing Countries	
Thailand	$221
Kazakhstan	$262
Colombia	$340
Malaysia	$361
Botswana	$379
China	$398
Mexico	$461
Turkey	$468

TABLE 7-6

Low-Income Developing Countries: Per Capital Health Expenditures

Low-Income Developing Countries	
Ethiopia	$27
Nepal	$45
India	$62
Nigeria	$79
Honduras	$199

China provides an interesting example of financing and service delivery development in middle-income developing countries. Financing reforms replaced fully publicly funded services with three new arrangements tied to work status and residence: (1) Urban Employee Basic Medical Insurance in 1998 (incorporating privately funded medical savings accounts—a concept pioneered in Singapore); (2) Urban Resident Basic Medical Insurance in 2007; and (3) New Rural Cooperative Medical Scheme in 2007. The urban employee program is an SHI model reflecting the rapid rate of economic growth and increasing incomes for urban workers. Starting in 2013, the Chinese government increasingly emphasized the development of new private hospitals and promotion of private insurance in urban areas. These and other health sector reforms became possible as continued strong economic growth over 30 years raised an estimated 300 million Chinese into the middle class, generating the requisite private as well as public revenues to underpin major structural health sector change.

Service delivery in developing countries varies widely in access, quality, and outcomes across and also within many developing countries. Medical services and tertiary institutions in urban areas of China, for example, operate at a substantially higher standard of service than those typically available in poorer rural regions. Similar disparities exist in wealthier parts of India such as Rajasthan, whereas in poorer states such as Bihar, primary care is mostly delivered by community "volunteers" with basic medical training, supervised by a GP

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

new providers and services. Second, available funding will need to be more effectively targeted on needed and appropriate services, with minimized managerial inefficiencies and substantially less political corruption.

available in poorer rural regions. Similar disparities exist in wealthier parts of India such as Rajasthan, whereas in poorer states such as Bihar, primary care is mostly delivered by community "volunteers" with basic medical training, supervised by a GP.

Two critical challenges for all developing country health systems are contingent on generating adequate future funding flows. First, the current push from United Nations agencies to achieve universal health coverage will require additional public and private sector funding to pay for the necessary new providers and services. Second, available funding will need to be more effectively targeted on needed and appropriate services, with minimized managerial inefficiencies and substantially less political corruption.

Both forms of expanded funding will be dependent on strong national and global economic growth, which in turn will require continued country-level economic and political reforms. Achieving both funding-related objectives will require considerable international as well as national effort.

FURTHER READING

Barber SL et al: *Price Setting and Price Regulation in Health Care: Lessons for Advancing Universal Health Coverage*. Geneva, World Health Organization, Organization for Economic Co-operation and Development, 2019. *https://apps.who.int/iris/bitstream/handle/10665/325547/9789241515924-eng.pdf*.

Figueras J, McKee M (eds): *Health Systems, Health, Wealth, and Societal Well-Being: Assessing the Case for Investing in Health Systems*. Maidenhead, Open University Press/McGraw-Hill Education, 2011. *www.euro.who.int/__data/assets/pdf_file/0007/164383/e96159.pdf.*

Haseltine W: *Affordable Excellence: The Singapore Health Story*. Washington, Brookings Institution Press, 2013. *www.brookings.edu/wp-www.brookings.edu/wp-content/uploads/2016/07/AffordableExcellencePDF.pdf*.

Kuhlmann E et al (eds): *The Palgrave International Handbook on Healthcare Policy and Governance*. London, Palgrave MacMillan, 2015.

Rice T et al: *United States of America: Health System Review*. Health in Transition (HiT) Series 15 (3). Brussels, European Observatory on Health Systems and Policies, 2013. *www.euro.who.int/__data/assets/pdf_file/0019/215155/HiT-United-States-of-America.pdf*.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 7: Global Diversity of Health System Financing and Delivery, Richard B. Saltman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 8: The Safety and Quality of Health Care

David W. Bates

INTRODUCTION

Safety and quality are two of the central dimensions of health care. In recent years, it has become easier to measure safety and quality, and it is increasingly clear that performance in both dimensions could be much better. The public is—with good justification—demanding measurement and accountability, and payment for services will increasingly be based on performance in these areas. Thus, physicians must learn about these two domains, how they can be improved, and the relative strengths and limitations of the current ability to measure them.

Safety and quality are closely related but do not completely overlap. The Institute of Medicine has suggested in a seminal series of reports that safety is the first part of quality and that the health care system must first and foremost guarantee that it will deliver safe care, although quality is also pivotal. In the end, it is likely that more net clinical benefit will be derived from improving quality than from improving safety, though both are important and safety is in many ways more tangible to the public. The first section of this chapter will address issues relating to the safety of care and the second will cover quality of care.

SAFETY IN HEALTH CARE

Safety Theory and Systems Theory

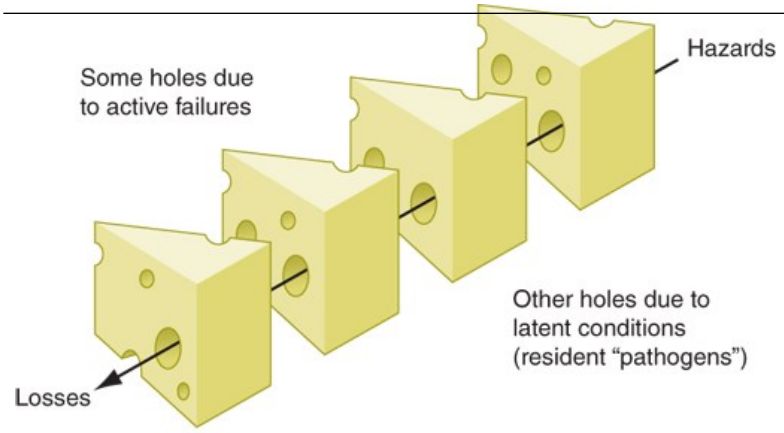
Safety theory clearly points out that individuals make errors all the time. Think of driving home from the hospital: you intend to stop and pick up a quart of milk on the way home but find yourself entering your driveway without realizing how you got there. Everybody uses low-level, semiautomatic behavior for many activities in daily life; this kind of error is called a *slip*. Slips occur often during care delivery—e.g., when people intend to write an order but forget because they must complete another action first. *Mistakes*, by contrast, are errors of a higher level; they occur in new or nonstereotypic situations in which conscious decisions are being made. An example would be dosing of a medication with which a physician is not familiar. The strategies used to prevent slips and mistakes are often different.

Systems theory suggests that most accidents occur as the result of a series of small failures that happen to line up in an individual instance so that an accident can occur **(Fig. 8-1)**. It also suggests that most individuals in an industry such as health care are trying to do the right thing (e.g., deliver safe care) and that most accidents thus result from defects in systems. Systems should be designed both to make errors less likely and to identify those that do inevitably occur.

FIGURE 8-1

"Swiss cheese" diagram. Reason argues that most accidents occur when a series of "latent failures" are present in a system and happen to line up in a given instance, resulting in an accident. Examples of latent failures in the case of a fall might be that the unit is unusually busy and the floor happens to be wet. *(Adapted from J Reason: BMJ 320:768, 2000.)*

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 8: The Safety and Quality of Health Care, David W. Bates
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Factors That Increase the Likelihood of Errors

Many factors ubiquitous in health care systems can increase the likelihood of errors, including fatigue, stress, interruptions, complexity, and transitions. The effects of fatigue in other industries are clear, but its effects in health care have been more controversial until recently. For example, the accident rate among truck drivers increases dramatically if they work over a certain number of hours in a week, especially with prolonged shifts. A recent study of house officers in the intensive care unit demonstrated that they were about one-third more likely to make errors when they were on a 24-h shift than when they were on a schedule that allowed them to sleep 8 h the previous night. The American College of Graduate Medical Education has moved to address this issue by putting in place the 80-h workweek. Although this stipulation is a step forward, it does not address the most important cause of fatigue-related errors: extended-duty shifts. High levels of stress and heavy workloads also can increase error rates. Thus, in extremely high-pressure situations, such as cardiac arrests, errors are more likely to occur. Strategies such as using protocols in these settings can be helpful, as can simple recognition that the situation is stressful.

Interruptions also increase the likelihood of error and occur frequently in health care delivery. It is common to forget to complete an action when one is interrupted partway through it by a page, for example. Approaches that may be helpful in this area include minimizing interruptions and setting up tools that help define the urgency of an interruption.

Complexity represents a key issue that contributes to errors. Providers are confronted by streams of data (e.g., laboratory tests and vital signs), many of which provide little useful information but some of which are important and require action or suggest a specific diagnosis. Tools that emphasize specific abnormalities or combinations of abnormalities may be helpful in this area.

Transitions between providers and settings are also common in health care, especially with the advent of the 80-h workweek, and generally represent points of vulnerability. Tools that provide structure in exchanging information—for example, when transferring care between providers—may be helpful.

The Frequency of Adverse Events in Health Care

Most large studies focusing on the frequency and consequences of adverse events have been performed in the inpatient setting; some data are available for nursing homes, but much less information is available about the outpatient setting. The Harvard Medical Practice Study, one of the largest studies to address this issue, was performed with hospitalized patients in New York. The primary outcome was the adverse event: an injury caused by medical management rather than by the patient's underlying disease. In this study, an event either resulted in death or disability at discharge or prolonged the length of hospital stay by at least 2 days. Key findings were that the adverse event rate was 3.7% and that 58% of the adverse events were considered preventable. Although New York is not representative of the United States as a whole, the study was replicated later in Colorado and Utah, where the rates were essentially similar. Since then, other studies using analogous methodologies have been performed in various developed nations, and the rates of adverse events in these countries appear to be ~10%. Rates of safety issues appear to be even higher in developing and transitional countries; thus, this is clearly an issue of global proportions.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 8: The Safety and Quality of Health Care, David W. Bates
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

5% were falls.

In the Harvard Medical Practice Study, adverse drug events (ADEs) were most common, accounting for 19% of all adverse events, and were followed in frequency by wound infections (14%) and technical complications (13%). Almost half of adverse events were associated with a surgical procedure. Among nonoperative events, 37% were ADEs, 15% were diagnostic mishaps, 14% were therapeutic mishaps, 13% were procedure-related mishaps, and 5% were falls.

ADEs have been studied more than any other error category. Studies focusing specifically on ADEs have found that they appear to be much more common than was suggested by the Harvard Medical Practice Study, although most other studies use more inclusive criteria. Detection approaches in the research setting include chart review and the use of a computerized ADE monitor, a tool that explores the database and identifies signals that suggest an ADE may have occurred. Studies that use multiple approaches find more ADEs than does any individual approach, and this discrepancy suggests that the true underlying rate in the population is higher than would be identified by a single approach. About 6–10% of patients admitted to U.S. hospitals experience an ADE.

Injuries caused by drugs are also common in the outpatient setting. One study found a rate of 21 ADEs per every 100 patients per year when patients were called to assess whether they had had a problem with one of their medications. The severity level was lower than in the inpatient setting, but approximately one-third of these ADEs were preventable.

The period immediately after a patient is discharged from the hospital appears to be very risky. A recent study of patients hospitalized on a medical service found an adverse event rate of 19%; about one-third of those events were preventable, and another one-third were ameliorable (i.e., they could have been made less severe). ADEs were the single leading error category.

Prevention Strategies

Most work on strategies to prevent adverse events has targeted specific types of events in the inpatient setting, with nosocomial infections and ADEs having received the most attention. Nosocomial infection rates have been reduced greatly in intensive care settings, especially by using checklists. For ADEs, several strategies have been found to reduce the medication error rate, although it has been harder to demonstrate that they reduce the ADE rate overall, and no studies with adequate power to show a clinically meaningful reduction have been published.

Implementation of checklists to ensure that specific actions are carried out has had a major impact on rates of catheter-associated bloodstream infection and ventilator-associated pneumonia, two of the most serious complications occurring in intensive care units. The checklist concept is based on the premise that several specific actions can reduce the frequency of these issues; when these actions are all taken for every patient, the result has been an extreme reduction in the frequency of the associated complication. These practices have been disseminated across wide areas in the state of Michigan.

Computerized physician order entry (CPOE) linked with clinical decision support reduces the rate of serious medication errors, defined as those that harm someone or have the potential to do so. In one study, CPOE, even with limited decision support, decreased the serious medication error rate by 55%. CPOE can prevent medication errors by suggesting a default dose, ensuring that all orders are complete (e.g., that they include dose, route, and frequency), and checking orders for allergies, drug–drug interactions, and drug–laboratory issues. In addition, clinical decision support can suggest the right dose for a patient, tailoring it to the level of renal function and age. In one study, patients with renal insufficiency received the appropriate dose only one-third of the time without decision support, whereas that fraction increased to approximately two-thirds with decision support; moreover, with such support, patients with renal insufficiency were discharged from the hospital half a day earlier. As of 2019, over 95% of U.S. hospitals had implemented CPOE, although the decision support often is still limited.

Another technology that can improve medication safety is bar coding linked with an electronic medication administration record. Bar coding can help ensure that the right patient gets the right medication at the right time. Electronic medication administration records can make it much easier to determine what medications a patient has received. Studies to assess the impact of bar coding on medication safety are under way, and the early results are promising. Another technology to improve medication safety is “smart pumps.” These pumps can be set according to which medication is being given and at what dose; the health care professional will receive a warning if too high a dose is about to be administered.

The National Safety Picture

Several organizations, including the National Quality Forum and The Joint Commission, have made recommendations for improving safety. The National Quality Forum has released recommendations to U.S. hospitals about what practices will most improve the safety of care, and all hospitals are

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 8: The Safety and Quality of Health Care, David W. Bates
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

70).

National Quality Forum has released recommendations to U.S. hospitals about what practices will most improve the safety of care, and all hospitals are expected to implement these recommendations. Many of these practices arise frequently in routine care. One example is “readback,” the practice of recording all verbal orders and immediately reading them back to the physician to verify the accuracy of what was heard. Another is the consistent use of standard abbreviations and dose designations; some abbreviations and dose designations are particularly prone to error (e.g., 7U may be read as 70).

Measurement of Safety

Measuring the safety of care is difficult and expensive, since adverse events are, fortunately, rare. Most hospitals rely on spontaneous reporting to identify errors and adverse events, but the sensitivity of this approach is very low, with only ~1 in 20 ADEs reported. Promising research techniques involve searching the electronic record for signals suggesting that an adverse event has occurred. These methods are not yet in wide use but will probably be used routinely in the future. Claims data have been used to identify the frequency of adverse events; this approach works much better for surgical care than for medical care and requires additional validation. The net result is that, except for a few specific types of events (e.g., falls and nosocomial infections), hospitals have little idea about the true frequency of safety issues.

Nonetheless, all providers have the responsibility to report problems with safety as they are identified. All hospitals have spontaneous reporting systems, and if providers report events as they occur, those events can serve as lessons for subsequent improvement.

Conclusions about Safety

It is abundantly clear that the safety of health care can be improved substantially. As more areas are studied closely, more problems are identified. Much more is known about the epidemiology of safety in the inpatient setting than in outpatient settings. A number of effective strategies for improving inpatient safety have been identified and are increasingly being applied. Some effective strategies are also available for the outpatient setting. Transitions appear to be especially risky. The solutions to improving care often entail the consistent use of systematic techniques such as checklists and often involve leveraging of information technology. Nevertheless, solutions will also include many other domains, such human factors techniques, team training, and a culture of safety.

QUALITY IN HEALTH CARE

Assessment of quality of care has remained somewhat elusive, although the tools for this purpose have increasingly improved. Selection of health care and measurement of its quality are components of a complex process.

Quality Theory

Donabedian has suggested that quality of care can be categorized by type of measurement into structure, process, and outcome. *Structure* refers to whether a particular characteristic is applicable in a particular setting—e.g., whether a hospital has a catheterization laboratory or whether a clinic uses an electronic health record. *Process* refers to the way care is delivered; examples of process measures are whether a Pap smear was performed at the recommended interval or whether an aspirin was given to a patient with a suspected myocardial infarction. *Outcome* refers to what happens—e.g., the mortality rate in myocardial infarction. It is important to note that good structure and process do not always result in a good outcome. For instance, a patient may present with a suspected myocardial infarction to an institution with a catheterization laboratory and receive recommended care, including aspirin, but still die because of the infarction.

Quality theory also suggests that overall quality will be improved more in the aggregate if the performance level of all providers is raised rather than if a few poor performers are identified and punished. This view suggests that systems changes are especially likely to be helpful in improving quality, since large numbers of providers may be affected simultaneously.

The theory of *continuous quality improvement* suggests that organizations should be evaluating the care they deliver on an ongoing basis and continually making small changes to improve their individual processes. This approach can be very powerful if embraced over time.

Several specific tools have been developed to help improve process performance. One of the most important is the Plan-Do-Check-Act cycle **(Fig. 8-2)**. This approach can be used for “rapid cycle” improvement of a process—e.g., the time that elapses between a diagnosis of pneumonia and administration of antibiotics to the patient. Some statistical tools, such as control charts, are often used in conjunction to determine whether progress is being made. Because most medical care includes one or many processes, this tool is especially important for improvement.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 8: The Safety and Quality of Health Care, David W. Bates

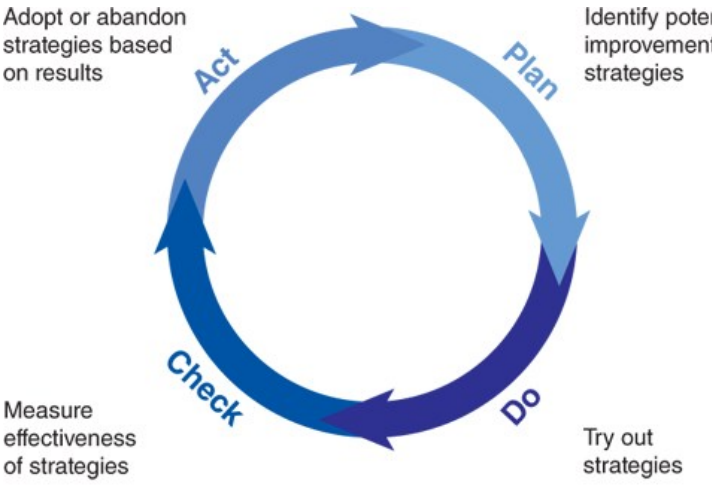
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Plan-Do-Check-Act cycle. This approach can be used to improve a specific process rapidly. First, planning is undertaken, and several potential improvement strategies are identified. Next, these strategies are evaluated in small “tests of change.” “Checking” entails measuring whether the

is being made. Because most medical care includes one or many processes, this tool is especially important for improvement.

FIGURE 8-2

Plan-Do-Check-Act cycle. This approach can be used to improve a specific process rapidly. First, planning is undertaken, and several potential improvement strategies are identified. Next, these strategies are evaluated in small "tests of change." "Checking" entails measuring whether the strategies have appeared to make a difference, and "acting" refers to acting on the results.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Factors Relating to Quality

Many factors can decrease the level of quality, including stress to providers, high or low levels of production pressure, and poor systems. Stress can have an adverse effect on quality because it can lead providers to omit important steps, as can a high level of production pressure. Low levels of production pressure sometimes can result in worse quality, as providers may be bored or have little experience with a specific problem. Poor systems can have a tremendous impact on quality, and even extremely dedicated providers typically cannot achieve high levels of performance if they are operating within a poor system.

Data about the Current State of Quality

A study published by the RAND Corporation in 2006 provided the most complete picture of quality of care delivered in the United States to date. The results were sobering. The authors found that, across a wide range of quality parameters, patients in the United States received only 55% of recommended care overall; there was little variation by subtype, with scores of 54% for preventive care, 54% for acute care, and 56% for care of chronic conditions. The authors concluded that, in broad terms, the chances of getting high-quality care in the United States were little better than those of winning a coin flip.

Work from the Dartmouth Atlas of Health Care evaluating geographic variation in use and quality of care demonstrates that, despite large variations in utilization, there is no positive correlation between the two variables at the regional level. An array of data demonstrate, however, that providers with larger volumes for specific conditions, especially for surgical conditions, do have better outcomes.

Strategies for Improving Quality and Performance

Many specific strategies can be used to improve quality at the individual level, including rationing, education, feedback, incentives, and penalties. *Rationing* has been effective in some specific areas, such as persuading physicians to prescribe within a formulary, but it generally has been resisted. *Education* is effective in the short run and is necessary for changing opinions, but its effect decays fairly rapidly with time. *Feedback* on performance can be given at either the group or the individual level. Feedback is most effective if it is individualized and is given in close temporal proximity to the original events. *Incentives* can be effective, and many believe that they will prove to be a key to improving quality, especially if pay-for-performance with sufficient incentives is broadly implemented (see below). *Penalties* produce provider resentment and are rarely used in health care.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 8: The Safety and Quality of Health Care, David W. Bates
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

be implemented together with a high degree of fidelity. Many hospitals have implemented a bundle for ventilator-associated pneumonia in the intensive care unit that includes five measures (e.g., ensuring that the head of the bed is elevated). These hospitals have been able to improve

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

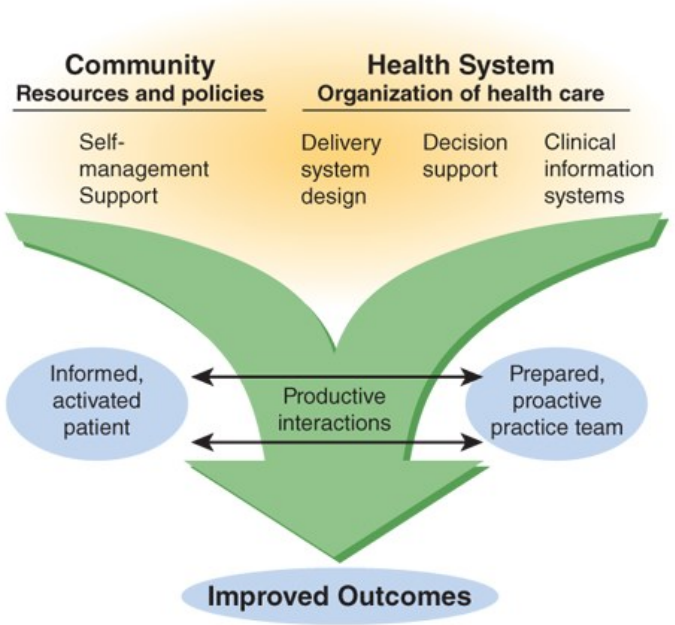
with sufficient incentives is broadly implemented (see below). *Penalties* produce provider resentment and are rarely used in health care.

Another set of strategies for improving quality involves changing the systems of care. An example would be introducing reminders about which specific actions need to be taken at a visit for a specific patient—a strategy that has been demonstrated to improve performance in certain situations, such as the delivery of preventive services. Another approach that has been effective is the development of "bundles" or groups of quality measures that can be implemented together with a high degree of fidelity. Many hospitals have implemented a bundle for ventilator-associated pneumonia in the intensive care unit that includes five measures (e.g., ensuring that the head of the bed is elevated). These hospitals have been able to improve performance substantially. Another technique is SCAMPs, or Standardized Clinical Assessment and Management Plans. These are care guidelines developed by clinicians who identify key steps in workflow and decisions to help improve the process outcomes.

Perhaps the most pressing need is to improve the quality of care for chronic diseases. The Chronic Care Model has been developed by Wagner and colleagues **(Fig. 8-3)**; it suggests that a combination of strategies is necessary (including self-management support, changes in delivery system design, decision support, and information systems) and that these strategies must be delivered by a practice team composed of several providers, not just a physician.

FIGURE 8-3

The Chronic Care Model, which focuses on improving care for chronic diseases, suggests that (1) delivery of high-quality care requires a range of strategies that must closely involve and engage the patient and (2) team care is essential. *(From EH Wagner et al: Eff Clin Pract 1:2, 1998.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Available evidence about the relative efficacy of strategies in reducing hemoglobin A_{1c} (HbA_{1c}) in outpatient diabetes care supports this general premise. It is especially notable that the outcome was the HbA_{1c} level, as it has generally been much more difficult to improve outcome measures than process measures (such as whether HbA_{1c} was measured). In this meta-analysis, a variety of strategies were effective, but the most effective ones were the use of team changes and the use of a case manager. When cost-effectiveness is considered in addition, it appears likely that an amalgam of strategies will be needed. However, the more expensive strategies, such as the use of case managers, probably will be implemented widely only if pay-for-performance takes hold.

The evidence linking better performance on quality metrics assessing process and outcomes varies greatly by condition. For example, there is strong evidence that performing Pap smears results in better outcomes in patients who develop cervical cancer, but the evidence for many other conditions is far more tenuous.

National State of Quality Measurement

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 8: The Safety and Quality of Health Care, David W. Bates
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

National Quality Forum, and the Agency for Healthcare Research and Quality. The data are housed at the Centers for Medicare and Medicaid Services, which publicly releases performance data on the measures on a website called *Hospital Compare* (*www.cms.gov/Medicare/Quality-Initiatives-Patient-*

In the inpatient setting, quality measurement is now being performed by a very large proportion of hospitals for several conditions, including myocardial infarction, congestive heart failure, pneumonia, and surgical infection prevention; 20 measures are included in all. This is the result of the Hospital Quality Initiative, which represents a collaboration among many entities, including the Hospital Quality Alliance, The Joint Commission, the National Quality Forum, and the Agency for Healthcare Research and Quality. The data are housed at the Centers for Medicare and Medicaid Services, which publicly releases performance data on the measures on a website called *Hospital Compare* (*www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalCompare.html*). These data are reported voluntarily and are available for a very high proportion of the nation's hospitals. Analyses demonstrate substantial regional variation in quality and important differences among hospitals. Analyses by The Joint Commission for similar indicators reveal that performance on measures by hospitals has improved over time and that, as might be hoped, lower performers have improved more than higher performers.

Public Reporting

Overall, public reporting of quality data is becoming increasingly common. There are now commercial websites that have quality-related data for most regions of the United States, and these data can be accessed for a fee. Similarly, national data for hospitals are available. The evidence to date indicates that patients have not made much use of such data, but that the data have had an important effect on provider and organization behavior. Instead, patients have relied on provider reputation to make choices, partly because little information was available until very recently and the information that was available was not necessarily presented in ways that were easy for patients to access. Problems still exist with quality metrics; many can be "gamed," and even though providers are now nearly universally using electronic health records (EHRs), most metrics come from claims that include many inaccuracies. More metrics that leverage EHRs are sorely needed. However, many authorities think that, as more information about quality becomes available, it will become increasingly central to patients' choices about where to access care.

Pay-for-Performance

Currently, providers in the United States get paid the same amount for a specific service, regardless of the quality of care delivered. The pay-for-performance theory suggests that, if providers are paid more for higher-quality care, they will invest in strategies that enable them to deliver that care. The current key issues in the pay-for-performance debate relate to (1) how effective it is, (2) what levels of incentives are needed, and (3) what perverse consequences are produced. The evidence on effectiveness is limited, although a number of studies are ongoing. With respect to incentive levels, most quality-based performance incentives have accounted for merely 1–2% of total payment in the United States to date. In the United Kingdom, however, 40% of general practitioners' salaries have been placed at risk according to performance across a wide array of parameters; this approach has been associated with substantial improvements in reported quality performance, although it is still unclear to what extent this change represents better performance versus better reporting. The potential for perverse consequences exists with any incentive scheme. One problem is that, if incentives are tied to outcomes, there may be a tendency to transfer the sickest patients to other providers and systems. Another concern is that providers will pay too much attention to quality measures with incentives and ignore the rest of the quality picture. The validity of these concerns remains to be determined. Nonetheless, it appears likely that, under health care reform, the use of various pay-for-performance schemes is likely to increase.

CONCLUSIONS

The safety and quality of care in the United States could be improved substantially. A number of available interventions have been shown to improve the safety of care and should be used more widely; others are undergoing evaluation or soon will be. Quality also could be dramatically better, and the science of quality improvement continues to mature. Implementation of value-based approaches such as accountable care that include pay-for-performance related to safety and quality should make it much easier for organizations to justify investments in improving safety and quality parameters, including health information technology. However, many improvements will also require changing the structure of care—e.g., moving to a more team-oriented approach and ensuring that patients are more involved in their own care. Payment reform focusing on value seems very likely to progress and will likely include both positive incentives and penalties related to safety and quality performance. Measures of safety are still relatively immature and could be made much more robust; it would be particularly useful if organizations had measures they could use in routine operations to assess safety at a reasonable cost, and substantial research is addressing this. Although the quality measures available are more robust than those for safety, they still cover a relatively small proportion of the entire domain of quality, and more measures need to be developed. The public and payers are demanding better information about safety and quality as well as better performance in these areas. The clear implication is that these domains will have to be addressed directly by providers.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 8: The Safety and Quality of Health Care, David W. Bates
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

1998. [PubMed: 9794308]

FURTHER READING

Bates DW et al: Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA 280:1311, 1998. [PubMed: 9794308]

Bates DW et al: Two decades since to err is human: An assessment of progress and emerging priorities in patient safety. Health Aff (Millwood) 37:1736, 2018. [PubMed: 30395508]

Berwick DM: Era 3 for medicine and health care. JAMA 315:1329, 2016. [PubMed: 26940610]

Brennan TA et al: Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med 324:370, 1991. [PubMed: 1987460]

Chertow GM et al: Guided medication dosing for inpatients with renal insufficiency. JAMA 286:2839, 2001. [PubMed: 11735759]

Institute of Medicine. Report: To err is human: Building a safer health system. 1999. *https://www-nap-edu.libproxy.umflint.edu/resource/9728/To-Err-is-Human-1999--report-brief.pdf.*

Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. 2001. *https://www-nap-edu.libproxy.umflint.edu/catalog/10027/crossing-the-quality-chasm-a-new-health-system-for-the*.

Landrigan C et al: Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med 351:1838, 2004. [PubMed: 15509817]

McGlynn EA et al: The quality of health care delivered to adults in the United States. N Engl J Med 348:2635, 2003. [PubMed: 12826639]

Pronovost P et al: An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 355:2725, 2006. Erratum in: N Engl J Med 356:2660, 2007. [PubMed: 17192537]

Starmer AJ et al: Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. JAMA 310:2262, 2013. [PubMed: 24302089]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 8: The Safety and Quality of Health Care, David W. Bates
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 9: Diagnosis: Reducing Errors and Improving Quality

Gordon Schiff

INTRODUCTION

Diagnosing patients' illnesses is the essence of medicine. Patients present to doctors seeking an answer to the question, "What is wrong with me?" Ideally, no clinician would want to treat a patient without knowing the diagnosis or, worse yet, erroneously treat a misdiagnosed illness. From the earliest moments of medical school, the defining quest toward becoming a knowledgeable and proficient physician is learning how to put a diagnostic label on patients' symptoms and physical findings, and clinicians pride themselves on being "good diagnosticians." Yet the centuries-old paradigm of mastering a long list of diseases, understanding their pathophysiology, and knowing the cardinal ways they manifest themselves in signs and symptoms, while still of fundamental importance, is being challenged by new insights illuminated by the glaring spotlight of diagnostic errors. Basic internal medicine diseases, such as asthma, pulmonary embolism, congestive heart failure, seizures, strokes, ruptured aneurysms, depression, and cancer, are misdiagnosed at shockingly high rates, often with 20–50% of patients either being mislabeled as having these conditions (false-positive diagnoses) or having their diagnosis missed or delayed (false negatives). How and why do physicians so often get it wrong, and what can we do to both diagnose and treat the problem of delayed diagnosis or misdiagnosis?

Diagnosis is both an ancient art and a modern science. The current science of diagnosis, however, goes far beyond what typically comes to clinicians' and patients' minds when they conjure up images of state-of-the-art molecular, genetic, or imaging technologies. Improvements in diagnosis are just as likely to come from other areas, many with origins outside of medicine, as they are from advanced diagnostic testing modalities. These diverse sciences that the field of diagnostic safety has, and must, draw from include systems and human factors engineering, reliability science, cognitive psychology, decision sciences, forensic science, clinical epidemiology, health services research, decision analysis, network medicine, learning health systems theory, medical sociology, team dynamics and communication, risk assessment and communication, information and knowledge management, and health information technology, especially artificial intelligence and clinical decision support. A clinician reading this chapter is likely to find this list of overlapping and intersecting domains quite daunting. However, rather than feeling overwhelmed, we urge readers to view them as the basic science supports that will ultimately make their lives easier and diagnosis more accurate and timely. Rather than feeling intimidated, clinicians should feel a sense of relief and assurance in understanding that good diagnosis does not rest entirely on their shoulders. Instead, it is a systems property, where an infrastructure and a team, one that especially includes the patient, can in a coordinated way work together to achieve more reliable and optimal diagnosis.

EMERGENCE OF DIAGNOSIS ERROR AS AN IMPORTANT PATIENT SAFETY ISSUE

Over the past decade, a series of studies culminating in a landmark report from the U.S. National Academy of Medicine (NAM), *Improving Diagnosis in Health Care,* have shone a spotlight on diagnostic errors. Reports from patient surveys, malpractice claims, and safety organizations, such as the ECRI and the National Patient Safety Foundation (now part of Institute for Healthcare Improvement), have found that diagnostic errors are the leading type of medical error. Although errors in diagnosis are defined in various ways, the NAM Committee defined diagnostic error as "the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the patient." One way to visualize diagnostic errors is through a Venn diagram **(Fig. 9-1)**, which illustrates the fact that many things can go wrong in the diagnostic process (e.g., failure to ask an important history question, physical examination sign overlooked, laboratory specimen erroneously switched between two patients, x-ray not followed up), but this usually does not result in a wrong diagnosis or patient harm. Similarly, a patient can be misdiagnosed but unharmed, without any identifiable error in the care received. Our greatest concern is where these three circles intersect, with conservative estimates suggesting that 40,000–80,000 patients die each year in U.S. hospitals alone from diagnostic errors. The NAM report outlined eight recommendations that are the foundation for this chapter **(Table 9-1)**.

FIGURE 9-1

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

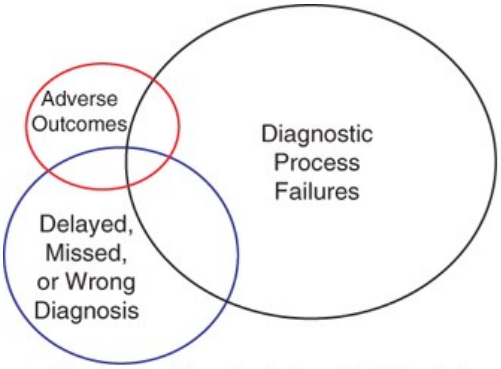
Leape. Acad Med 87:133, 2012.)

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

FIGURE 9-1

What is a diagnosis error? *(Adapted from GD Schiff et al: Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project, in Advances in Patient Safety: from Research to Implementation. Vol. 2 Concepts and Methodology, Rockville, MD, 2005, pp 255-278, and GD Schiff, L Leape: Acad Med 87:135, 2012.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 9-1

National Academy of Medicine Recommendations for Improving Diagnosis in Health Care

1. Facilitate more effective teamwork in the diagnostic process among health care professionals, patients, and their families.
2. Enhance professional education and training in the diagnostic process in areas such as clinical reasoning; teamwork; communication with patients, families, and other health care professionals; and appropriate use of diagnostic tests.
3. Ensure that health information technologies support patients and health care professionals in the diagnostic process.
4. Develop and deploy approaches to identify, learn from, and reduce diagnostic errors and near misses in clinical practice including providing systematic feedback on diagnostic performance.
5. Establish a work system and culture that supports the diagnostic process and improvements in diagnostic performance.
6. Develop a reporting environment and medical liability system that facilitates improved diagnosis by learning from diagnostic errors and near misses.
7. Design a payment and care delivery environment that supports the diagnostic process.
8. Provide dedicated funding for research on the diagnostic process and diagnostic errors.

NEW WAYS TO THINK ABOUT DIAGNOSIS AND DIAGNOSTIC ERRORS

Medical textbooks have historically given attention to "clinician reasoning" and associated cognitive heuristics and biases. Errors in clinical reasoning can be summarized in three broad groups: (1) hasty judgments, (2) biased judgments, and (3) inaccurate probability estimates. Research from cognitive psychology has identified scores of common mental shortcuts or "heuristics" humans are prone to use in everyday life, many of which are useful for efficient diagnosis but can also lead to biases and errors. **Table 9-2** lists some of the common cognitive biases that can lead diagnosis astray (this topic is discussed further in **Chap. 4**).

TABLE 9-2

Selected Cognitive Biases Contributing to Diagnostic Errors

1. Premature closure: accepting a diagnosis before it has been fully verified
2. Anchoring: tendency to fixate on a specific symptom or piece of information early in the diagnostic process with subsequent failure to appropriately adjust
3. Confirmation bias: tendency to look for confirming evidence to support one's diagnostic hypothesis, rather than disconfirming evidence to refute it
4. Search satisficing: tendency to call off a search, satisfied once a piece of data or presumed explanation is found, and not considering/searching for

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

disease in a low-prevalence population using a test with 5% false-positive rate)
7. Knowledge deficit (on part of provider, with accompanying lack of awareness)

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

TABLE 9-2

Selected Cognitive Biases Contributing to Diagnostic Errors

1. Premature closure: accepting a diagnosis before it has been fully verified
2. Anchoring: tendency to fixate on a specific symptom or piece of information early in the diagnostic process with subsequent failure to appropriately adjust
3. Confirmation bias: tendency to look for confirming evidence to support one's diagnostic hypothesis, rather than disconfirming evidence to refute it
4. Search satisficing: tendency to call off a search, satisfied once a piece of data or presumed explanation is found, and not considering/searching for additional findings or diagnoses
5. Availability bias: tendency to give too much weight to diagnoses that come more readily to mind (e.g., recent dramatic case)
6. Base-rate neglect: failing to adequately take into account prevalence of a particular disease (e.g., erroneously interpreting a positive test as indicating disease in a low-prevalence population using a test with 5% false-positive rate)
7. Knowledge deficit (on part of provider, with accompanying lack of awareness)
8. Framing bias: Judgement overly influenced by the way the problem was presented (how it was framed in words, settings, situations)
9. Social/demographic/stereotype bias: biases from personal or cultural beliefs about women, minorities, or other patient groups for whom prejudices may distort diagnostic assessment

However, clinicians will also benefit from having a better understanding of diagnosis as a "system" rather than just what takes place in clinicians' minds. Classic teaching exhorting trainees and practicing physicians to have a broad differential and "high index of suspicion" for various diseases is challenged not only by these unconscious biases but also by limitations of human memory, information shortfalls, constrained encounter time, system process failures, and the myriad nonspecific symptoms that patients bring to clinicians. Many symptoms are self-limited, defy a precise diagnosis or etiology, and do not portend harmful outcomes. Insights from safety and cognitive sciences call for rethinking traditional approaches to diagnosis and suggest new approaches to overcome current limitations **(Table 9-3)**.

TABLE 9-3

New Models for Conceptualizing Diagnosis and Diagnosis Improvement

TRADITIONAL WAYS OF THINKING ABOUT DIAGNOSIS AND DIAGNOSTIC ERROR	NEW PARADIGMS/BETTER WAYS TO THINK ABOUT DIAGNOSIS AND IMPROVING DIAGNOSIS
General	
A good diagnostician gets it right the first time, almost all of the time	Diagnosis is an inexact science with inherent uncertainties Goal is to minimize errors and delays via more reliable systems and follow-up
Lore of masterful/skillful academic expert diagnostician who knows/recalls everything; need to look to them if seeking diagnostic excellence	Less reliance on (fallible) human memory Quality diagnosis is based on well-coordinated distributed network/team of people and reliable processes All patients entitled to receive quality diagnosis, regardless of where and from whom they receive care
Diagnosis is the doctor's job	Co-production of diagnosis among clinicians (including lab, radiology, specialists, nurses, social workers) and, especially, the patient and family
Patients often viewed as overly anxious, exaggerating, time-consuming, questioning, with sometimes unreasonable demands and expectations	Patients are key allies in diagnosis; hold key information Need to address understandable/legitimate fears, desires for explanations Leveraging patient questions and questioning of diagnosis to stimulate rethinking the diagnosis where needed
	treatment

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

social workers) and, especially, the patient and family	
demands and expectations	Leveraging patient questions and questioning of diagnosis to stimulate rethinking the diagnosis where needed
Diagnosis and treatment as separate stages in patient care (i.e., make a diagnosis, then treat)	Prioritizing diagnostic efforts to target treatable conditions More integrated strategies and timing for testing and treatment depending on urgency for treatment
Clinical practices	
Order lots of tests to avoid missing diagnoses	Judicious ordering: targeted, well-organized data and testing Appreciation of test limitations (false positive or negative, incidental findings, overdiagnosis, test risks) and resulting harms
More referrals to avoid missing rarer/specialized diagnoses; concomitant utilization barriers (copays, prior authorization) to minimize overuse	"Pull systems" to lower barriers and make it easier to pose questions, obtain real-time virtual consults Co-management approaches to enable collaborative watch-and-wait conservative strategies where appropriate
Frequent empirical drug trials when uncertain of diagnosis	Conservative use of drugs to avoid confusing clinical picture or labeling patients with diseases they may not have
Physician attention/efforts to ensure disease screening	Automating, delegating clerical functions; teamwork to free up physician cognitive time
Diagnosis errors and challenges	
Diagnostic error viewed as a personal failing Errors classified as either "system" or "cognitive"	Many errors/delays rooted in processes and system design/failures Errors multifactorial with interwoven, interacting, and inseparable cognitive and system factors
Errors are infrequent; hit-and-miss ways to learn about errors	Errors are common; systematic proactive follow-up is needed to recognize potential for errors Surveilling of high-risk situations and one's own diagnostic performance and outcomes
Clinicians' reactions: denial, defensive, others to blame, pointing to others also making similar errors	Culture of actively and nondefensively seeking to uncover, dig deep to learn from, and share errors and lessons
Dreading complex, frustrating diagnostic dilemmas	Welcoming/enjoying intellectual/professional challenges Adequate support (time, help, consultations) for more complex patients
Diagnoses as distinct labels, events	Diagnoses can be indistinct, interacting comorbidities, socially constructed, multifactorial, evolving over time, or have overlapping genotype-phenotype expressions
Documentation/communication	
Viewed as time-consuming, mindless, primarily to document for billing code and/or bulwark against malpractice claims	Documentation as useful tool for reflecting, crafting, sharing assessments, differential diagnosis, reflecting about unanswered questions Opportunities for decision support interacting with computer Notes open for patients to read to help understand and critique diagnosis
Say and write as little as possible about uncertainties, lest it be used against you in malpractice allegation	Share uncertainties to maximize communication and engagement with other caregivers, patients
Don't let patient know about errors so they don't become	Patients have right to honest disclosure; often find out about errors anyway (e.g., cancer
(test results: "We'll call if anything is abnormal.")	is doing, monitor outcomes

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Notes open for patients to read to help understand and critique diagnosis	
Don't let patient know about errors so they don't become angry, mistrustful, or sue	Patients have right to honest disclosure; often find out about errors anyway (e.g., cancer evolves); anticipate, engage their concerns
Patients advised to call if not better; no news is good news (test results: "We'll call if anything is abnormal.")	Systematic proactive follow-up to close loop on tests and symptoms, to check how patient is doing, monitor outcomes
Global remedies	
Knowing/memorizing more medical knowledge	Knowing more about the patient (including psychosocial, past history, environmental contexts)
Attention to the "objective" data (physical exam, tests) to reliably make diagnoses	Renewed emphasis on history, history-taking, listening Acknowledgement of ubiquitous subjective cognitive biases; efforts to anticipate, recognize, counteract
Exhortations to have "high index of suspicion" of various diagnoses	Less reliance on memory recall of lectures/reading; more just-in-time info look-up Affordances, alerts to red flags engineered into workflow Delineation of "don't miss" diagnoses with design of context-relevant decision support reminders
Ensuring physician is copied on everything, thorough/voluminous notes, widespread reminders/alerts	Biggest problem is no longer lack of access to information, but rather information overload; strategies to organize, minimize
Continuing medical education (CME) courses to expand medical knowledge	Real-time, context-aware reminders of pitfalls, critical differential diagnoses, and key differentiating features. Ready access to medical references, second opinions
Redundancies, double-checks	Recognition that single, highly reliable systems are often better than multiple halfway solutions. Clear delineation of responsibilities for follow-up tasks
Fear of malpractice suits to motivate physicians to be more careful and practice defensive medicine	Drive out fear, making it safe to learn from and share errors Shared situational awareness of where pitfalls lurk
More accountability, financial incentives, and penalties tied to performance metrics	Clinician engagement in improvement based on trust, collaboration, professionalism, financial neutrality Metric modesty, recognizing many best practices yet to be defined/proven
More rules, requirements; target outliers for better compliance	Standardization with flexibility; learning from deviations
More time with patients	Better time spent with patients: offloading distractions, more efficient history collection/organization, longitudinal continuity, and, where needed, additional time to talk/think/explain during, before, or after visits Easier access for patients to reach or be seen by clinicians when experiencing symptoms
Reflex changes in response to errors	Avoiding "tampering," which entails understanding/diagnosing difference between "special cause" versus "common cause" (random) variation

Source: Modified from GD Schiff: Quality and Safety in Health Care 2013.

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Given variations and overlap in ways patients present, illnesses evolve, and tests perform, it is often not possible or practical to "make" a definitive diagnosis, particularly in the primary care setting early in the course of a patient's illness. Clinicians need to harness these uncertainties to both

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Source: Modified from GD Schiff: Quality and Safety in Health Care 2013.

UNCERTAINTY IN DIAGNOSIS

Given variations and overlap in ways patients present, illnesses evolve, and tests perform, it is often not possible or practical to "make" a definitive diagnosis, particularly in the primary care setting early in the course of a patient's illness. Clinicians need to harness these uncertainties to both engineer situational awareness of where things can go wrong and create safety nets to protect patients against harms from delayed diagnosis and misdiagnosis. Terms such as *preliminary diagnosis*, *working diagnosis*, *differential diagnosis*, *deferred diagnosis*, *undiagnosed illness*, *diagnoses with uncertain or multifactorial etiologies*, *intermittent diagnoses*, *multiple/dual diagnoses*, *self-diagnosis*, or at times *contested diagnosis* need to be part of our vocabulary, thinking, and communications with patients to convey that diagnosis is often imprecise. Anxious patients worried about a condition, for example, cancer, COVID-19 infection, or a diagnosis to which a relative or a friend has recently succumbed, come seeking reassurance and may not welcome an uncertain answer. Thus, we have to work with patients, listen to and respect their concerns, and take their symptoms seriously yet modestly acknowledge our limitations. We need to tailor this approach to patients' differing levels of health literacy, trust in our clinical advice, and experiences with the health system.

DON'T MISS DIAGNOSES AND RED FLAGS

Uncertainty should not be a license for complacency. Particularly for diseases that (1) progress rapidly, (2) require specific treatments that depend on making the correct diagnosis, or (3) have public health or contagion implications, clinicians need to be poised, and systems designed, to consider and, where appropriate, pursue critical "don't miss" diagnoses. While clinicians are generally aware of more common "don't miss" diagnoses (e.g., acute myocardial infarction, sepsis), **Table 9-4** illustrates examples of less common diagnoses that warrant similar consideration. Throughout this textbook, readers should orient themselves to recognize such critical diagnoses and think about presentations and syndromes where they may be lurking.

TABLE 9-4

Examples of "Don't Miss" Diagnoses

INFECTIONS/INFLAMMATION	CARDIAC/ISCHEMIC/BLEEDING	METABOLIC/HEMATOLOGIC/ENVIRONMENTAL
Spinal epidural abscess	Aortic dissection Leaking/ruptured abdominal aortic aneurysm	Diabetes ketoacidosis Hyperosmolar hyperglycemia
Necrotizing fasciitis	Pericardial tamponade	Myxedema/thyrotoxicosis
Meningitis	Wolff-Parkinson-White Prolonged QT	Addison's disease
Endocarditis	Pulmonary embolism	B_{12} deficiency anemia
Peritonsillar abscess	Tension pneumothorax	von Willebrand's disease
Tuberculosis-active pulmonary, other	Acute mesenteric ischemia Sigmoid volvulus	Hemochromatosis
COVID-19 infection	Esophageal, bowel perforation	Celiac sprue
Guillain-Barré syndrome	Cerebellar hemorrhage	Carbon monoxide poisoning
Ebola infection	Spinal cord compression	Food poisoning
Temporal arteritis	Testicular, ovarian torsion	Malignant hyperthermia
Angioedema	Retroperitoneal hemorrhage	Tumor lysis syndrome Hypo-/hypercalcemia

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 9-4

Examples of "Don't Miss" Diagnoses

INFECTIONS/INFLAMMATION	CARDIAC/ISCHEMIC/BLEEDING	METABOLIC/HEMATOLOGIC/ENVIRONMENTAL
Spinal epidural abscess	Aortic dissection Leaking/ruptured abdominal aortic aneurysm	Diabetes ketoacidosis Hyperosmolar hyperglycemia
Necrotizing fasciitis	Pericardial tamponade	Myxedema/thyrotoxicosis
Meningitis	Wolff-Parkinson-White Prolonged QT	Addison's disease
Endocarditis	Pulmonary embolism	B_{12} deficiency anemia
Peritonsillar abscess	Tension pneumothorax	von Willebrand's disease
Tuberculosis-active pulmonary, other	Acute mesenteric ischemia Sigmoid volvulus	Hemochromatosis
COVID-19 infection	Esophageal, bowel perforation	Celiac sprue
Guillain-Barré syndrome	Cerebellar hemorrhage	Carbon monoxide poisoning
Ebola infection	Spinal cord compression	Food poisoning
Temporal arteritis	Testicular, ovarian torsion	Malignant hyperthermia
Rhabdomyolysis	Ectopic pregnancy	Alcohol, benzodiazepine, barbiturate withdrawal
Angioedema	Retroperitoneal hemorrhage	Tumor lysis syndrome Hypo-/hypercalcemia

An important related concept is so-called "red flags" or "alarm symptoms." This construct has its origins in guidelines for back pain but has increasingly been applied to many other problems, such as headache, red eye, swollen joint, or even abdominal pain and chest pain. Examples of widely cited red flags for back pain that should trigger consideration of more serious etiologies include fever, weight loss, history of malignancy or intravenous drug use, or neurologic signs and symptoms. In theory, many presenting syndromes could benefit from identification of such clues to more serious diagnoses. Evidence-based medicine calls for better data on the sensitivity, specificity, yield, and discriminatory ability of various clinical "red flag" clues; yet, few have been rigorously evaluated. Nonetheless, clinicians find them useful as simple ways to reassure themselves and their patients that a common symptom such as back pain or headache is, or is not, likely an indicator of more urgent or serious pathology.

Interwoven with the challenges of not missing critical diagnoses is the problem of overtesting and overdiagnosis—performing unnecessary and even potentially harmful tests whose benefit does not justify the risks or costs or that may lead to diagnoses that would have never caused any symptoms or problems. Thoughtful diagnosticians need to weigh carefully this "other side of the coin" of missed diagnosis to avoid such harms and expenses.

DIAGNOSTIC PITFALLS

One of the important ways of learning in medicine is learning from the missteps of those who have walked the path ahead of us. By learning about commonly missed diagnoses and the ways accurate, timely diagnosis went astray, we can avoid making similar mistakes. Anticipating the potential for similar types of errors can both create situational awareness of traps to avoid and contribute to learning from our own personal and collective patterns

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

assessment proceeding all the way to invasive biopsy, if necessary. Diagnostic pitfalls can be classified into a number of generic scenarios (Table 9-5). We now have large databases that have the potential to track "diagnoses outcomes"—i.e., whether a new diagnosis emerges that suggests an initial

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

commonly missed diagnoses and the ways accurate, timely diagnosis went astray, we can avoid making similar mistakes. Anticipating the potential for similar types of errors can both create situational awareness of traps to avoid and contribute to learning from our own personal and collective patterns of mistakes. Several studies have examined common or recurring pitfalls in diagnosis. An example of a common disease-specific diagnostic pitfall in breast cancer diagnosis is ordering a mammogram for a woman with a palpable breast lump and, when the mammogram returns as normal, reassuring her that cancer has been "ruled out" by the negative test. Any mass or lesion palpable on physical examination probably needs more careful assessment proceeding all the way to invasive biopsy, if necessary. Diagnostic pitfalls can be classified into a number of generic scenarios **(Table 9-5)**. We now have large databases that have the potential to track "diagnoses outcomes"—i.e., whether a new diagnosis emerges that suggests an initial diagnosis was incorrect or a diagnosis of a patient's symptoms was suboptimally delayed. This should, in the future, allow us to more rigorously focus on these cases, to identify contributing factors and recurring patterns, and to help point the way for systemwide improvement strategies.

TABLE 9-5

Generic Types of Diagnostic Pitfalls

PITFALL	EXAMPLES
Disease A mistaken for disease B Diseases often mistaken/misdiagnosed with each other	• Aortic dissection misdiagnosed as acute myocardial infarction • Bipolar disorder misdiagnosed as depression
Misinterpretation of test result(s) False-positive or false-negative results with failure to recognize test limitations	• Breast lump dismissed after negative mammogram • Negative COVID-19 test early or late in course
Failure to recognize atypical presentation, signs, and symptoms	• Apathetic hyperthyroidism • Sepsis in elderly patient who is afebrile or hypothermic
Failure to assess appropriately the urgency of diagnosis Urgency of the clinical situation was not appreciated and/or delays critical diagnoses	• Compartment syndrome • Pericardial tamponade • Tension pneumothorax
Perils of intermittent symptoms or misleading evolution Intermittent symptoms dismissed due to normal findings (exam, lab, electrocardiogram) when initially seen	• "Lucid interval" in traumatic epidural hematoma • Paroxysmal arrhythmias • Intermittent hydrocephalus (Bruns' syndrome)
Confusion arising from response/masking by empiric treatment	• Empiric treatment with steroids, proton pump inhibitors, antibiotics, pain medication erroneously masking serious diagnosis
Chronic disease or comorbidity presumed to account for new symptoms Especially in medically complex patients	• Septic joint signs misattributed to chronic rheumatoid arthritis • Mental status change due to infection or medication misattributed to underlying dementia
Rare diagnosis: failure to consider or know	• Many; fortunately, by definition, rare, but still warrant consideration especially if urgent or treatable
Drug or environmental factor not considered/overlooked Underlying etiology causing/contributing to symptoms, or disease progression not sought, uncovered	• Ventricular arrhythmia related to QT-prolonging drug • Achilles tendon rupture related to quinolone drugs
Failure to appreciate limitations of physical exam Now with ↑ telemedicine, missing physical exam entirely	• Overweighing absence of tenderness, swelling in deep vein thrombosis • Missing pill-rolling tremor during telemedicine visit

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

on these cases, to identify contributing factors and recurring patterns, and to help point the way for systematic improvement strategies.

TABLE 9-5

Generic Types of Diagnostic Pitfalls

PITFALL	EXAMPLES
Disease A mistaken for disease B Diseases often mistaken/misdiagnosed with each other	• Aortic dissection misdiagnosed as acute myocardial infarction • Bipolar disorder misdiagnosed as depression
Misinterpretation of test result(s) False-positive or false-negative results with failure to recognize test limitations	• Breast lump dismissed after negative mammogram • Negative COVID-19 test early or late in course
Failure to recognize atypical presentation, signs, and symptoms	• Apathetic hyperthyroidism • Sepsis in elderly patient who is afebrile or hypothermic
Failure to assess appropriately the urgency of diagnosis Urgency of the clinical situation was not appreciated and/or delays critical diagnoses	• Compartment syndrome • Pericardial tamponade • Tension pneumothorax
Perils of intermittent symptoms or misleading evolution Intermittent symptoms dismissed due to normal findings (exam, lab, electrocardiogram) when initially seen	• "Lucid interval" in traumatic epidural hematoma • Paroxysmal arrhythmias • Intermittent hydrocephalus (Bruns' syndrome)
Confusion arising from response/masking by empiric treatment	• Empiric treatment with steroids, proton pump inhibitors, antibiotics, pain medication erroneously masking serious diagnosis
Chronic disease or comorbidity presumed to account for new symptoms Especially in medically complex patients	• Septic joint signs misattributed to chronic rheumatoid arthritis • Mental status change due to infection or medication misattributed to underlying dementia
Rare diagnosis: failure to consider or know	• Many; fortunately, by definition, rare, but still warrant consideration especially if urgent or treatable
Drug or environmental factor not considered/overlooked Underlying etiology causing/contributing to symptoms, or disease progression not sought, uncovered	• Ventricular arrhythmia related to QT-prolonging drug • Achilles tendon rupture related to quinolone drugs
Failure to appreciate risk factors for particular disease	• Family history of breast, colorectal cancer not solicited and/or weighed in diagnostic evaluation or screening
Failure to appreciate limitations of physical exam Now with ↑ telemedicine, missing physical exam entirely	• Overweighing absence of tenderness, swelling in deep vein thrombosis • Missing pill-rolling tremor during telemedicine visit

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

healthy safety culture that will allow it to grow and flourish. While clinicians may be inclined to view "safety culture" as something too subjective to be important in their quest to make a definitive diagnosis, this view is misguided. Multiple studies have demonstrated adverse consequences resulting

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

DIAGNOSIS SAFETY CULTURE

Just as diagnosing bacterial infections relies on a proper culture medium to grow and identify etiologic organisms, good diagnosis also requires a healthy safety culture that will allow it to grow and flourish. While clinicians may be inclined to view "safety culture" as something too subjective to be important in their quest to make a definitive diagnosis, this view is misguided. Multiple studies have demonstrated adverse consequences resulting from organizational cultures that inhibit openness, learning, and sharing and create a climate where staff and patients are afraid to speak up when they observe problems or have questions. Most importantly, patients need to be encouraged to question diagnoses and be heard, particularly when they are not responding to treatment as expected or developing symptoms that are either not consistent with the diagnosis or represent possible red flags for other diagnoses or complications.

Studies examining "high-reliability organizations" outside of medicine and "learning health care organizations" have distilled a series of fundamental properties that are correlated with more reliable and safer outcomes. Just as a thermometer or recording of a pulse can suggest how ill a patient is, we now have instruments that can measure safety culture. These safety measurement tools typically are validated staff surveys that assess (1) communication about errors with staff willingness to report mistakes because they do not feel these mistakes are held against them; (2) openness and encouragement to talk about hospital/office problems; (3) existence of a learning culture that seeks to learn from errors and improve based on lessons learned; (4) leadership commitment to safety, prioritizing safety over production speed and the "bottom line" by providing adequate staffing and resources to operate safely; and (5) accountability and transparency for following up safety events and concerns. Each of these generic culture attributes translates into specific implications for diagnostic safety. These include the following:

- Making it "safe" for clinicians to admit and share diagnostic errors
- Proactive identification, ownership, and accountability regarding error-prone diagnostic workflow processes (particularly around test results, referrals, and patient follow-up)
- Leadership making diagnosis improvement a top priority based on recognition that patients and malpractice insurers report that diagnostic errors are the leading patient safety problem
- Mutual trust and respect for challenges that clinicians often face in making diagnoses and caution in applying the lens of hindsight bias in judging what in retrospect might seem like an "obvious" diagnosis that a clinician initially missed

HEALTH INFORMATION TECHNOLOGY AND THE FUTURE OF DIAGNOSIS

Clinicians now spend more time interacting with computers than they do interacting with patients. This is especially true for diagnosis and will likely be even more so in the future. Interactions with patients, consultants, and other staff are increasingly mediated through the computer. Key activities, such as collecting patients' history (past and current), interpreting data to make a diagnosis, conveying diagnostic assessments (to others on the team and, increasingly, to the patient via open notes), and tracking diagnostic trajectories as they evolve over time, are now computer based. With the rise of telemedicine, even elements of the physical examination have been rerouted to electronic encounters.

While many complain the computer has "gotten in the way" of good diagnosis, distracting clinicians from quality time listening to patients and miring doctors in reading and writing notes filled with copied/pasted/templated information of questionable currency and accuracy, medicine needs to harness the computer's capabilities to improve diagnosis **(Table 9-6)**. Although these basic diagnosis-supporting capabilities should be the foundation of the design of health information technology and everyday workflow, electronic medical records have historically been largely designed around other needs, such as ordering medications and billing and malpractice documentation. They need to be radically redesigned to better support diagnostic processes, as well as save, rather than squander, clinicians' time.

TABLE 9-6

Areas Where Health Information Technology Has Potential to Help Improve Diagnosis and Reduce Errors

FUNCTION	EXAMPLES
Facilitate collection/gathering of	• Quickly access past history from prior care at same and outside institutions
Enhanced information entry, organization, and display	• Visually enhanced flowsheets showing trends, relationships to treatment • Reorganized notes to facilitate summarization and simplification and prevent items from getting lost

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

diagnostic processes, as well as save, rather than squander, clinicians' time.

TABLE 9-6

Areas Where Health Information Technology Has Potential to Help Improve Diagnosis and Reduce Errors

FUNCTION	EXAMPLES
Facilitate collection/gathering of information	• Quickly access past history from prior care at same and outside institutions • Electronic collection of history of present illness, review of systems, and social determinant risks in advance of visits
Enhanced information entry, organization, and display	• Visually enhanced flowsheets showing trends, relationships to treatment • Reorganized notes to facilitate summarization and simplification and prevent items from getting lost
Generating differential diagnosis	• Automated creation of lists of diagnoses to consider based on patient's symptoms, demographics, risks
Weighing diagnoses likelihoods	• Tools to assist in calculation of posttest (Bayesian) probabilities
Aids for formulating diagnostic plan, intelligent test ordering	• Entering a diagnostic consideration (e.g., celiac disease, pheochromocytoma) and computer suggests most appropriate diagnostic test(s) and how to order
Access to diagnostic reference information	• Info-buttons instantly linking symptom or diagnosis relevant questions to Harrison's, Up-to-Date chapters, references
Ensuring more reliable follow-up	• Hardwiring "closed loops" to ensure abnormal labs, missed referrals, worrisome symptoms are tracked and followed up
Support screening for early detection	• Collaborative tools that patients, clinicians, and offices can use to know when due, order and track screening based on individualized demographics, risk factors, prior tests
Collaborative diagnosis; access to specialist	• Real-time posing/answering of questions • Electronic consults; virtual co-management
Facilitating feedback on diagnoses	• Feeding back new diagnoses (from downstream providers, patients) that emerge suggesting potential misdiagnosis/errors to clinicians, ERs who saw patient previously

Abbreviation: ERs, emergency rooms.

Source: Modified from G Schiff, DW Bates: N Engl J Med 362:1066, 2010, and R El-Karah et al: BMJ Qual Saf Suppl 2:ii40, 2013.

DIAGNOSIS OF DIAGNOSIS ERRORS AND SAFETY: PRACTICAL CONCLUSIONS

In practice, there are frequent and meaningful opportunities for improving diagnosis in each of the three NAM-defined areas to make it a) more reliable, b) timely, and c) to improve diagnosis-related communication with patients. Clinicians in training, practicing physicians, nurses, and others should develop the habit of regularly asking themselves three questions about individual patients in their care, and another three questions regarding

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

2. What doesn't fit? (making sure unexplained abnormal findings are not dismissed)

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

should develop the habit of regularly asking themselves three questions about individual patients in their care, and another three questions regarding the systems in which they work. For each patient being assessed, clinicians should ask:

1. What else might this be? (forcing a differential diagnosis to be made)
2. What doesn't fit? (making sure unexplained abnormal findings are not dismissed)
3. What critical diagnoses are important not to miss? (injecting consideration of "don't miss" diagnoses, red flags, and known pitfalls)

and to diagnose safely, each practitioner must recognize that he or she is working within a larger system. Questions to be asking continually, ensuring we are maximizing reliability and timeliness and minimizing potential for errors, include:

1. Do we have reliable "closed loop" systems to provide reliable, ideally automated tracking and following up of patients' symptoms, abnormal laboratory or imaging findings, and critical referrals that we order?
2. What is the culture-of-safety climate in our organization, office, or clinic?
3. How does the electronic (or even paper) medical record as currently implemented help versus impair efficient, timely, accurate, and fail-safe diagnosis, and how can it be improved?

To take these questions to the next stage, an international movement dedicated to studying and improving diagnosis has emerged. These efforts include annual conferences of clinicians, researchers, and patients; the formation of the Society for Improving Diagnosis in Medicine (SIDM); and convening of a broad coalition of organizations, including the American Board of Internal Medicine (ABIM), the American College of Physicians (ACP), and the Society of Hospital Medicine (SHM), committed to increasing awareness and action. Ultimately, collectively tackling the challenges of improving the quality of diagnosis will transform the way clinicians and patients work together to co-produce better diagnoses.

FURTHER READING

Gandhi TK, Singh H: Reducing the risk of diagnostic error in the COVID-19 era. J Hosp Med 15:363, 2020. [PubMed: 32490798]

Graber ML et al: The impact of electronic health records on diagnosis. Diagnosis (Berl) 4:211, 2017. [PubMed: 29536944]

National Academies of Sciences, Engineering, and Medicine. 2015. *Improving Diagnosis in Health Care. https://doi-org.libproxy.umflint.edu/10.17226/21794.* Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, DC.

Newman-Toker DE et al: Serious misdiagnosis-related harms in malpractice claims: The "big three"—Vascular events, infections, and cancers. Diagnosis (Berl) 6:227, 2019. [PubMed: 31535832]

Schiff GD et al: Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project, in *Advances in Patient Safety: From Research to Implementation. Vol 2: Concepts and Methodology*. Rockville, MD, Agency for Healthcare Research and Quality, 2005.

Schiff GD et al: Ten principles for more conservative, care-full diagnosis. Ann Intern Med 169:643, 2018. [PubMed: 30285046]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 9: Diagnosis: Reducing Errors and Improving Quality, Gordon Schiff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 11: Ethical Issues in Clinical Medicine

Christine Grady; Bernard Lo

INTRODUCTION

Physicians face novel ethical dilemmas that can be perplexing and emotionally draining. For example, telemedicine, artificial intelligence, handheld personal devices, and learning health care systems all hold the promise of more coordinated and comprehensive care, but also raise concerns about confidentiality, the doctor–patient relationship, and responsibility. This chapter presents approaches and principles that physicians can use to address important vexing ethical issues they encounter in their work. Physicians make ethical judgments about clinical situations every day. They should prepare for lifelong learning about ethical issues so they can respond appropriately. Traditional professional codes and ethical principles provide instructive guidance for physicians but need to be interpreted and applied to each situation. When facing or struggling with a challenging ethical issue, physicians may need to reevaluate their basic convictions, tolerate uncertainty, and maintain their integrity while respecting the opinions of others. Physicians should articulate their concerns and reasoning, discuss and listen to the views of others involved in the case, and utilize available resources, including other health care team members, palliative care, social work, and spiritual care. Moreover, ethics consultation services or a hospital ethics committee can help to clarify issues and identify strategies for resolution, including improving communication and dealing with strong or conflicting emotions. Through these efforts, physicians can gain deeper insight into the ethical issues they face and usually reach mutually acceptable resolutions to complex problems.

APPROACHES TO ETHICAL PROBLEMS

Several approaches are useful for resolving ethical issues, including approaches based on ethical principles, virtue ethics, professional oaths, and personal values. These various sources of guidance may seem to conflict in a particular case, leaving the physician in a quandary. In a diverse society, different individuals may turn to different sources of moral guidance. In addition, general moral precepts often need to be interpreted and applied to a particular clinical situation.

ETHICAL PRINCIPLES

Ethical principles can serve as general guidelines to help physicians determine the right thing to do.

Respecting Patients

Physicians should always treat patients with respect, which entails understanding patients' goals, providing information, communicating effectively, obtaining informed and voluntary consent, respecting informed refusals, and protecting confidentiality. Different clinical goals and approaches are often feasible, and interventions can result in both benefit and harm. Individuals differ in how they value health and medical care and how they weigh the benefits and risks of medical interventions. Generally, physicians should respect patients' values and informed choices. Treating patients with respect is especially important when patients are responding to experiences of, or fears about, disrespect and discrimination.

GOALS AND TREATMENT DECISIONS

Physicians should provide relevant and accurate information for patients about diagnoses, current clinical circumstances, expected future course, prognosis, treatment options, and uncertainties, and discuss patients' goals of care. Physicians may be tempted to withhold a serious diagnosis, misrepresent it by using ambiguous terms, or limit discussions of prognosis or risks for fear that patients will become anxious or depressed. Providing honest information about clinical situations promotes patients' autonomy and trust as well as sound communication with patients and colleagues. When physicians have to share bad news with patients, they should adjust the pace of disclosure, offer empathy and hope, provide emotional support, and call on other resources such as spiritual care or social work to help patients cope. Some patients may choose not to receive such information or

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 11: Ethical Issues in Clinical Medicine, Christine Grady; Bernard Lo

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Physicians should engage their patients in shared decision-making about their health and their care, whenever appropriate. Physicians should discuss

and call on other resources such as spiritual care or social work to help patients cope. Some patients may choose not to receive such information or may ask surrogates to make decisions on their behalf, as is common with serious diagnoses in some traditional cultures.

SHARED DECISION-MAKING AND OBTAINING INFORMED CONSENT

Physicians should engage their patients in shared decision-making about their health and their care, whenever appropriate. Physicians should discuss with patients the nature, risks, and benefits of proposed care; any alternative; and the likely consequences of each option. Physicians promote shared decision-making by informing and educating patients, answering their questions, checking that they understand key issues, making recommendations, and helping them to deliberate. Medical jargon, needlessly complicated explanations, or the provision of too much information at once may overwhelm patients. Increasingly, decision aids can assist patients in playing a more active role in decision-making, improving the accuracy of their perception of risk and benefit, and helping them feel better informed and clearer about their values. Informed consent is more than obtaining signatures on consent forms and involves disclosure of honest and understandable information to promote understanding and choice. Competent, informed patients may refuse recommended interventions and choose among reasonable alternatives. In an emergency, treatment can be given without informed consent if patients cannot give their own consent and delaying treatment while surrogates are contacted would jeopardize patients' lives or health. People are presumed to want such emergency care unless they have previously indicated otherwise.

Respect for patients does not entitle patients to insist on any care or treatment that they want. Physicians are not obligated to provide interventions that have no physiologic rationale, that have already failed, or that are contrary to evidence-based practice recommendations or good clinical judgment. Public policies and laws also dictate certain decisions—e.g., allocation of scarce medical resources during a public health crisis such as the COVID-19 pandemic, use of cadaveric organs for transplantation, and requests for physician aid in dying.

CARING FOR PATIENTS WHO LACK DECISION-MAKING CAPACITY

Some patients are unable to make informed decisions because of unconsciousness, advanced dementia, delirium, or other medical conditions. Courts have the legal authority to determine that a patient is legally incompetent, but in practice, physicians usually determine when patients lack the capacity to make particular health care decisions and arrange for authorized surrogates to make decisions, without involving the courts. Patients with decision-making capacity can express a choice and appreciate their medical situation; the nature, risks, and benefits of proposed care; and the consequences of each alternative. Patient choices should be consistent with their values and not the result of delusions, hallucinations, or misinformation. Physicians should use available and validated assessment tools, resources such as psychiatry or ethics consultation, and clinical judgment to ascertain whether individuals have the capacity to make decisions for themselves. Patients should not be assumed to lack capacity if they disagree with recommendations or refuse treatment. Such decisions should be probed, however, to ensure the patient is not deciding based on misunderstandings and has the capacity to make an informed decision. When impairments are fluctuating or reversible, decisions should be postponed if possible until the patient recovers decision-making capacity.

When a patient lacks decision-making capacity, physicians seek an appropriate surrogate. Patients may designate a health care proxy through an advance directive or on a Physician Orders for Life-Sustaining Treatment form; such choices should be respected (see **Chap. 12**). For patients who lack decision-making capacity and have not previously designated a health care proxy, family members usually serve as surrogates. Statutes in most U.S. states delineate a prioritized list of relatives to make medical decisions. Patients' values, goals, and previously expressed preferences guide surrogate decisions. However, the patient's current best interests may sometimes justify overriding earlier preferences if an intervention is likely to provide significant benefit, previous statements do not fit the situation well, or the patient gave the surrogate leeway in decisions.

MAINTAINING CONFIDENTIALITY

Maintaining confidentiality is essential to respecting patients' autonomy and privacy; it encourages patients to seek treatment and to discuss problems candidly. However, confidentiality may be overridden to prevent serious harm to third parties or the patient. Exceptions to confidentiality are justified when the risk to others is serious and probable, no less restrictive measures can avert risk, and the adverse effects of overriding confidentiality are minimized and deemed acceptable by society. For example, laws require physicians to report cases of tuberculosis, sexually transmitted infection, elder or child abuse, and domestic violence.

Beneficence or Acting in Patients' Best Interests

The principle of *beneficence* requires physicians to act for the patient's benefit. Patients typically lack medical expertise, and illness may make them vulnerable. Patients rely on and trust physicians to treat them with compassion and provide sound recommendations and treatments aimed to

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 11: Ethical Issues in Clinical Medicine, Christine Grady; Bernard Lo

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

providing known ineffective interventions or acting without due care. Although often cited, this precept alone provides limited guidance because many beneficial interventions also pose serious risks.

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

vulnerable. Patients rely on and trust physicians to treat them with compassion and provide sound recommendations and treatments aimed to promote their well-being. Physicians encourage such trust and have a fiduciary duty to act in the best interests of patients, which should prevail over physicians' self-interest or the interests of third parties such as hospitals or insurers. A principle related to beneficence, "first do no harm," obliges physicians to prevent unnecessary harm by recommending interventions that maximize benefit and minimize harm and forbids physicians from providing known ineffective interventions or acting without due care. Although often cited, this precept alone provides limited guidance because many beneficial interventions also pose serious risks.

Physicians increasingly provide care within interdisciplinary teams and rely on consultation with or referral to specialists. Team members and consultants contribute different types of expertise to the provision of comprehensive, high-quality care for patients. Physicians should collaborate with and respect the contributions of the various interdisciplinary team members and should initiate and participate in regular communication and planning to avoid diffusion of responsibility and ensure accountability for quality patient care.

INFLUENCES ON PATIENTS' BEST INTERESTS

Conflicts arise when patients' refusal or request of interventions thwarts their own goals for care, causes serious harm, or conflicts with their best medical interests. For example, simply accepting a young asthmatic adult's refusal of mechanical ventilation for reversible respiratory failure, in the name of respecting autonomy, is morally constricted. Physicians should elicit patients' expectations and concerns, correct their misunderstandings, and try to persuade them to accept beneficial therapies. If disagreements persist after such efforts, physicians should call on institutional resources for assistance, but patients' informed choices and views of their own best interests should prevail.

Drug prices and out-of-pocket expenses for patients have been escalating in many parts of the world and may compromise care that is in the patients' best interests. Physicians should recognize that patients, especially those with high copayments or inadequate insurance, may not be able to afford prescribed tests and interventions. Physicians should strive to prescribe medications that are affordable and acceptable to the patient. Knowing what kind of insurance, if any, the patient has and whether certain medications are likely to be covered may help in determining appropriate prescriptions. Available alternatives should be considered and discussed. Physicians should follow up with patients who don't fill prescriptions, don't take their medications, or skip doses to explore whether cost and affordability are obstacles. It may be reasonable for physicians to advocate for coverage of nonformulary products for sound reasons, such as when the formulary drugs are less effective or not tolerated or are too costly for the patient to pay for out of pocket. These should be shared decisions with the patient to the extent possible.

Organizational policies and workplace conditions may sometimes conflict with patients' best interests. Physicians' focus and dedication to the well-being and interests of patients may be negatively influenced by perceived or actual staffing inadequacies, unfair wages, infrastructural deficiencies or lack of equipment, work-hour limitations, corporate culture, and threats to personal security in the workplace. Physicians should work with institutional leaders to ensure that policies and practices support their ability to provide quality care focused on patients' best interests.

Patients' interests are served by improvements in overall quality of care and the increasing use of evidence-based practice guidelines and performance benchmarking. However, practice guideline recommendations may not serve the interests of each individual patient, especially when another plan of care may provide substantially greater benefits. In prioritizing their duty to act in the patient's best interests, physicians should be familiar with relevant practice guidelines, be able to recognize situations that might justify exceptions, and advocate for reasonable exceptions.

Acting Justly

The principle of *justice* provides guidance to physicians about how to ethically treat patients and make decisions about allocating important resources, including their own time. *Justice* in a general sense means fairness: people should receive what they deserve. In addition, it is important to act consistently in cases that are similar in ethically relevant ways, in order to avoid arbitrary, biased, and unfair decisions. Justice forbids discrimination in health care based on race, religion, gender, sexual orientation, disability, age, or other personal characteristics **(Chap. 10)**.

ALLOCATION OF RESOURCES

Justice also requires fair allocation of limited health care resources. Universal access to medically needed health care remains an unrealized moral aspiration in the United States and many countries around the world. Patients with no or inadequate health insurance often cannot afford health care and lack access to safety-net services. Even among insured patients, insurers may deny coverage for interventions recommended by their physician. In this situation, physicians should advocate for patients' affordable access to indicated care, try to help patients obtain needed care, and work with institutions and policies to promote wider access. Doctors might consider—or patients might request—the use of lies or deception to obtain such benefits, for example, signing a disability form for a patient who does not meet disability criteria. Although motivated by a desire to help the patient

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 11: Ethical Issues in Clinical Medicine, Christine Grady; Bernard Lo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

to the concerns of those affected, and proportionate to the situation, including the supply relative to the need. In the 2019–2020 SARS-CoV-2/COVID-19 pandemic, some epicenters anticipated or faced shortages of staff, protective equipment, hospital and critical care beds, and ventilators, even after

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

institutions and policies to promote wider access. Doctors might consider—or patients might request—the use of lies or deception to obtain such benefits, for example, signing a disability form for a patient who does not meet disability criteria. Although motivated by a desire to help the patient, such deception breaches basic ethical guidelines and undermines physicians' credibility and trustworthiness.

Allocation of health care resources is unavoidable when resources are limited. Allocation policies should be fair, transparent, accountable, responsive to the concerns of those affected, and proportionate to the situation, including the supply relative to the need. In the 2019–2020 SARS-CoV-2/COVID-19 pandemic, some epicenters anticipated or faced shortages of staff, protective equipment, hospital and critical care beds, and ventilators, even after increasing supplies and modifying usual clinical procedures. Many jurisdictions developed guidelines for implementing crisis standards of care to allocate limited interventions and services. Under crisis standards of care, some aspects of conventional care are not possible and interventions may not be provided to all who might benefit or wish to receive them. Crisis standards of care aim to promote the good of the community by saving the most lives in the short term, using evidence-based criteria.

When demand for medications or other interventions exceeds the supply, allocation should be fair, strive to avoid discrimination, and mitigate health disparities. First-come, first-served allocation is not fair, because it disadvantages patients who experience barriers to accessing care. To avoid discrimination, allocation decisions should not consider personal social characteristics such as race, gender, or disability, nor consider insurance status or wealth. Allocation policies also should aspire to reduce health care disparities. U.S. African-American, Latino-American, and Native-American patients suffered a disproportionate number of COVID-19 cases and deaths, likely due in part to being employed in jobs that cannot be done remotely or with physical distancing, crowded housing, lack of health benefits, and poor access to health care.

Fair and well-considered guidelines help mitigate any emotional and moral distress that clinicians may experience making difficult allocation decisions. Authorizing triage officers or committees to make allocation decisions according to policies determined with public input allows treating physicians and nurses to dedicate their efforts to their patients. Ad hoc resource allocation by physicians at the bedside may be inconsistent, unfair, and ineffective. At the bedside, physicians should act as patient advocates within constraints set by society, reasonable insurance policies, and evidence-based practice. Many allocation decisions are made at the level of public policy, with physician and public input. For example, the United Network for Organ Sharing (*www.unos.org*) provides criteria for allocating scarce organs.

VIRTUE ETHICS

Virtue ethics focuses on physicians' character and qualities, with the expectation that doctors will cultivate virtues such as compassion, trustworthiness, intellectual honesty, humility, and integrity. Proponents argue that, if such characteristics become ingrained, they help guide physicians in unforeseen situations. Moreover, following ethical precepts or principles without any of these virtues could lead to uncaring doctor–patient relationships.

PROFESSIONAL OATHS AND CODES

Professional oaths and codes are useful guides for physicians. Most physicians take oaths during their medical training, and many are members of professional societies that have professional codes. Physicians pledge to the public and to their patients that they will be guided by the principles and values in these oaths or codes and commit to the spirit of the ethical ideals and precepts represented in oaths and professional codes of ethics.

PERSONAL VALUES

Personal values, cultural traditions, and religious beliefs are important sources of personal morality that help physicians address ethical issues and cope with any moral distress they may experience in practice. While essential, personal morality alone is a limited ethical guide in clinical practice. Physicians have role-specific ethical obligations that go beyond their obligations as good people, including the duties to obtain informed consent and maintain confidentiality discussed earlier. Furthermore, in a culturally and religiously diverse world, physicians should expect that some patients and colleagues will have personal moral beliefs that differ from their own.

ETHICALLY COMPLEX PROFESSIONAL ISSUES FOR PHYSICIANS

CLAIMS OF CONSCIENCE

Some physicians, based on their personal values, have conscientious objections to providing, or referring patients for, certain treatments such as contraception or physician aid in dying. Although physicians should not be asked to violate deeply held moral beliefs or religious convictions, patients

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

health care institution should be notified in advance of any conscientious objections to the provision of specific interventions. Since insurance often constrains patients' selection of physicians or health care facilities, switching providers can be burdensome. There are also important limits on claims

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

contraception or physician aid in dying. Although physicians should not be asked to violate deeply held moral beliefs or religious convictions, patients need medically appropriate, timely care and should always be treated with respect. Institutions such as clinics and hospitals have a collective ethical duty to provide care that patients need while making reasonable attempts to accommodate health care workers' conscientious objections—for example, when possible by arranging for another professional to provide the service in question. Patients seeking a relationship with a doctor or health care institution should be notified in advance of any conscientious objections to the provision of specific interventions. Since insurance often constrains patients' selection of physicians or health care facilities, switching providers can be burdensome. There are also important limits on claims of conscience. Health care workers may not insist that patients receive unwanted medical interventions. They also may not refuse to treat or discriminate against patients because of their race, ethnicity, disability, genetic information, or diagnosis. Such discrimination is illegal and violates physicians' duties to respect patients. Refusal to treat patients for other reasons such as sexual orientation, gender identity, or other personal characteristics is legally more controversial, yet ethically inappropriate because it falls short of helping patients in need and respecting them as persons.

PHYSICIAN AS GATEKEEPER

In some cases, patients may ask their physicians to facilitate access to services that the physician has ethical qualms about providing. For example, a patient might request a prescription for a cognitively enhancing medication to temporarily augment his cognitive abilities in order to take an exam or apply for employment. Patients may request more pain medication than the physician believes is warranted for the given situation or marijuana to facilitate sleep. Patients may ask their physician to sign a waiver to avoid vaccines for reasons that are not included in state exceptions (see **Chap. 3**). A physician may feel uncomfortable prescribing attention-deficit/hyperactivity disorder medications to a young child because she is not convinced that the possible benefit justifies the risks to the child despite the parent's request. In these circumstances, the physician should work with the patient or parent to understand the reasons for their requests, some of which might be legitimate. In addition to considering possible risks and benefits to the patient, the physician should consider how meeting the request might affect other patients, societal values, and public trust in the medical profession. If the physician determines that fulfilling the request requires deception, is unfair, jeopardizes her professional responsibilities, or is inconsistent with the patient's best medical interests, the physician should decline and explain the reasons to the patient.

MORAL DISTRESS

Health care providers, including residents, medical students, and experienced physicians, may experience moral distress when they feel that ethically appropriate action is hindered by institutional policies or culture, decision-making hierarchies, limited resources, or other reasons. Moral distress can lead to anger, anxiety, depression, frustration, fatigue, work dissatisfaction, and burnout. A physician's health and well-being can affect how he or she cares for patients. Discussing complex or unfamiliar clinical situations with colleagues and seeking assistance with difficult decisions can help alleviate moral distress, as can a healthy work environment characterized by open communication, mutual respect, and emphasis on the common goal of good patient care. In addition, physicians should take good care of their own well-being and be aware of the personal and system factors associated with stress, burnout, and depression. Health care organizations should provide a supportive work environment, counseling, and other support services when needed.

OCCUPATIONAL RISKS AND BURDENS

Physicians accept some physical risk in fulfilling their professional responsibilities, including exposure to infectious agents or toxic substances, violence in the workplace, and musculoskeletal injury. Nonetheless, most physicians, nurses, and other hospital staff willingly care for patients, despite personal risk and fear, grueling hours, and sometimes inadequate personal protective equipment or information. During the COVID-19 pandemic, many communities honored clinicians' dedication to professional ideals, and some medical students who were relieved from in-person patient care responsibilities volunteered to support front-line workers in other ways. The burdens of navigating professional and personal responsibilities fall more heavily on women health care providers. Health care institutions are responsible for reducing occupational risk and burden by providing proper information, training and supervision, protective equipment, infrastructure and workflow modifications, and emotional and psychological support to physicians. Clinical leaders need to acknowledge fears about personal safety and take steps to mitigate the impact of work on family responsibilities, moral distress, and burnout.

USE OF SOCIAL MEDIA AND PATIENT PORTALS

Increasingly, physicians use social and electronic media to share information and advice with patients and other providers. Social networking may be especially useful in reaching young or otherwise hard-to-access patients. Patients increasingly access their physicians' notes through patient portals,

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

harmful consequences for patients, physicians, or their colleagues. Unprofessional or careless posts that express frustration or anger over work incidents, disparage patients or colleagues, use offensive or discriminatory language, or reveal inappropriate personal information about the

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

especially useful in reaching young or otherwise hard-to-access patients. Patients increasingly access their physicians' notes through patient portals, which aim to transparently share information, promote patient engagement, and increase adherence. Physicians should be professional and respectful and consider patient confidentiality, professional boundaries, and therapeutic relationships when posting to social media or writing notes for the portal. Overall, appropriate use of these platforms can enhance communication and transparency while avoiding misunderstandings or harmful consequences for patients, physicians, or their colleagues. Unprofessional or careless posts that express frustration or anger over work incidents, disparage patients or colleagues, use offensive or discriminatory language, or reveal inappropriate personal information about the physician can have negative consequences. Physicians should separate professional from personal websites and accounts and follow institutional and professional society guidelines when communicating with patients.

CONFLICTS OF INTEREST

Acting in patients' best interests may sometimes conflict with a physician's self-interest or the interests of third parties such as insurers or hospitals. From an ethical viewpoint, patients' interests are paramount. Transparency, appropriate disclosure, and management of conflicts of interest are essential to maintain the trust of colleagues and the public. Disclosure requirements vary for different purposes, and software has been developed to assist physicians in complying with specific requirements. Importantly, not all conflicts are financial. Physicians sometimes face conflicts of commitment between their patient's interests and their own personal interests, professional goals, responsibilities, and aspirations. As mentioned earlier, physicians should prioritize patients' interests while recognizing possible conflicts and using disclosure, discussion with the chief of service, and management of the conflict or recusal when appropriate.

In addition to individual physicians, medical institutions may have conflicts of interest arising from patent rights, industry-funded research programs, and donations from individuals and companies. Institutions need to be transparent about the presence and amount of such relationships and make clear the steps taken to prevent such relationships from having an impact on clinical or financial decisions. If there is good evidence that a donor acted in ways that breached ethical or legal standards, the institution should take steps not to benefit from the donation or honor the donor.

FINANCIAL INCENTIVES

Physicians have financial incentives to improve the quality or efficiency of care that might lead some to avoid patients who are older, are chronically ill, or have more complicated problems, or to focus on benchmarked outcomes even when not in the best interests of individual patients. In contrast, fee-for-service payments might encourage physicians to order more interventions than necessary or to refer patients to laboratory, imaging, or surgical facilities in which they have a financial stake. Regardless of financial incentives, physicians should recommend available care that is in the patient's best interests—no more and no less.

RELATIONSHIPS WITH PHARMACEUTICAL COMPANIES

Financial relationships between physicians and industry are increasingly scrutinized. Many academic medical centers have banned drug-company gifts, including branded pens and notepads and meals to physicians, to reduce inappropriate risk of undue influence or subconscious feelings of reciprocity and to decrease possible influences on public trust or the costs of health care.

The federal Open Payments website provides public information on the payments and amounts that drug and device companies give to individual physicians by name. The challenge is to distinguish payments for scientific consulting and research contracts—which should be encouraged as consistent with professional and academic missions—from those for promotional speaking and consulting whose goal is to increase sales of company products.

LEARNING CLINICAL SKILLS

Medical students', residents', and physicians' interests in learning, which fosters the long-term goal of benefiting future patients, may sometimes conflict with the short-term goal of providing optimal care to current patients. When trainees are learning procedures on patients, they lack the proficiency of experienced physicians, and patients may experience inconvenience, discomfort, longer procedures, or increased risk. Increasingly, institutions are developing clinical skills laboratories for simulation-based medical education and requiring students to demonstrate proficiency before carrying out procedures such as venipuncture and intravenous lines in patients. Furthermore, teaching hospitals are establishing proceduralist services in which procedure-specialist faculty members directly supervise interns for procedures such as lumbar puncture and thoracentesis and certify their proficiency. Medical students may need to defer learning such invasive procedures until internship. Seeking patients' consent for trainee participation in their care is always important and is particularly important for intimate examinations, such as pelvic, rectal, breast, and testicular

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 11: Ethical Issues in Clinical Medicine, Christine Grady; Bernard Lo

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

difficult for patients to make informed choices about their care. Most patients, when informed, allow trainees to play an active role in their care.

RESPONSE TO MEDICAL ERRORS

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

certify their proficiency. Medical students may need to defer learning such invasive procedures until internship. Seeking patients' consent for trainee participation in their care is always important and is particularly important for intimate examinations, such as pelvic, rectal, breast, and testicular examinations, and for invasive procedures. Patients should be told who is providing care and how trainees are supervised. Failing to introduce students or not telling patients that trainees will be performing procedures undermines trust, may lead to more elaborate deception, and makes it difficult for patients to make informed choices about their care. Most patients, when informed, allow trainees to play an active role in their care.

RESPONSE TO MEDICAL ERRORS

Errors are inevitable in clinical medicine, and some errors cause harm to patients. Most errors are caused by lapses of attention or flaws in the system of delivering health care; only a small number result from blameworthy individual behavior. Many health care institutions have adopted a just culture system, which encourages open and honest reporting of errors as essential to quality learning and shifts the focus from individual blame to system design for improvement in quality and safety **(Chap. 8)**. This approach is more likely than a punitive approach to improve patient safety. However, professional discipline is appropriate for cases of gross incompetence, reckless behavior, physician impairment, and boundary violations. Physicians and students may fear that disclosing errors will damage their careers. Physicians and health care institutions show respect for patients by disclosing and explaining errors, offering an apology, offering appropriate compensation for harm done, and using errors as opportunities to improve the quality of care.

PHYSICIAN IMPAIRMENT

Physicians may hesitate to intervene when colleagues impaired by alcohol, drugs, or psychiatric or medical illness place patients at risk. However, society relies on physicians to regulate themselves. Colleagues of an impaired physician should take steps to protect patients and help their impaired colleague, starting with reporting their concerns to their clinical supervisor or director.

ETHICAL ISSUES IN CLINICAL RESEARCH

Clinical research is essential to translate scientific discoveries into beneficial interventions for patients. However, clinical research raises ethical concerns because participants face inconvenience and risks in research designed to advance scientific knowledge and not specifically to benefit them. Ethical guidelines require researchers to rigorously design and conduct research, minimize risk to participants, and obtain informed and voluntary consent from participants and approval from an institutional review board (IRB). IRBs determine that risks to participants are acceptable and have been minimized and recommend appropriate additional protections when research includes vulnerable participants.

Physicians may be clinical research investigators themselves or may be in a position to refer or recommend clinical trial participation to their patients. Physician-investigators are likely to feel some inherent tension between conducting research and providing health care. Awareness of this tension, familiarity with research ethics, collaboration with research and clinical team members, and utilizing research ethics consultation can help to mitigate tensions. Before starting clinical research, investigators should complete training in the ethics of clinical research, which is widely available.

Physicians also should be critical consumers of clinical research results and keep up with research advances that change standards of practice. Precision medicine initiatives aim to individualize clinical care by combining clinical information from electronic health records, genomic sequencing, and data from personal mobile devices. Furthermore, physicians and health care institutions are analyzing data routinely collected and available in electronic health records, leftover clinical specimens, and administrative data. Such studies encompass traditional discovery research as well as quality improvement, comparative effectiveness research, and learning health care systems. Efforts to improve the quality of care in real-world clinical settings are important but also raise new issues about informed consent, privacy, and risk.

EMERGING TECHNOLOGIES

Scientific advances in genome sequencing, gene editing (e.g., with CRISPR-Cas9), machine learning, artificial intelligence, computer–brain interfaces, and other technologies offer great promise for research and clinical care with the ultimate goal of improving the prediction, prevention, and treatment of disease. Groundbreaking innovations that have strong scientific plausibility need to be evaluated in rigorous clinical studies for efficacy and safety.

Physicians should keep up to date on the status of novel and often complex technologies as research evolves, data emerge, and technologies are incorporated into clinical practice. They can help their patients understand research findings and the evidence for clinical use, correct any misunderstandings, facilitate shared decision-making, and advocate for fair access to such therapies. Further, physicians should engage in professional and public discussion related to allocation of resources and fair access to expensive new therapies and emerging technologies and their

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 11: Ethical Issues in Clinical Medicine, Christine Grady; Bernard Lo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

(Chap. 69), are approved for use in several serious hematologic cancers, and gene therapies have been approved as safe and effective for clinical use in certain serious inherited diseases and cancers. Patients may request these and other complex, highly technical, and expensive therapies for

misunderstandings, facilitate shared decision-making, and advocate for fair access to such therapies. Further, physicians should engage in professional and public discussion related to allocation of resources and fair access to expensive new therapies and emerging technologies and their impact on overall health care affordability.

Certain cell-based therapies, such as peripheral blood stem cell transplantation **(Chap. 114)** and chimeric antigen receptor (CAR)-T cell therapy **(Chap. 69)**, are approved for use in several serious hematologic cancers, and gene therapies have been approved as safe and effective for clinical use in certain serious inherited diseases and cancers. Patients may request these and other complex, highly technical, and expensive therapies for unproven indications. Yet, claims of cures through unproven stem cell or gene-based "therapies" pose significant health and financial risks to patients without evidence of benefit. Physicians should help patients distinguish approved therapies from unproven claims and refer interested patients to well-designed clinical trials.

Medical applications of CRISPR-Cas9 are promising, and their safety and efficacy for particular clinical conditions are being carefully evaluated in clinical trials. Applications of CRISPR genome editing in somatic cells to modify or correct problematic genes could lay the foundation for treating a variety of serious diseases, including blood disorders, HIV, cancer, and hereditary blindness. Germline gene editing in blastocysts or embryos raises many ethical questions and is currently not permitted in the United States in clinical trials or clinical practice.

In artificial intelligence (AI), computers carry out tasks typically done by humans. Machine learning (ML) is a type of AI that automatically learns and improves its performance without explicit programming. Clinical algorithms using AI and ML can make diagnoses from radiology images, retinal scans, or skin photographs and identify patients at increased risk for surgical complications, critical care, or hospital readmission. However, such algorithms can also pose risks. Bias may occur if an algorithm was derived or validated from a data set in which groups who suffer from health disparities or poor health outcomes are underrepresented or if the algorithm predicts outcomes that are not clinically meaningful. To address these ethical concerns, researchers should assess AI algorithms in well-designed randomized clinical trials with clinical endpoints. Institutions should integrate validated and unbiased algorithms into clinical workflow without unduly burdening physicians and nurses and should check effectiveness and safety in their particular settings and patient populations.

Physicians should stay informed of emerging evidence about such technologies and the ethical challenges that accompany their use and always keep their patients' best interests and preferences at the forefront.

GLOBAL CONSIDERATIONS

INTERNATIONAL RESEARCH

Clinical research is often conducted across multiple sites and across national borders. Societal, legal, and cultural norms and perspectives about research may vary, and there are many ethical challenges. Physician-investigators involved in international research should be familiar with international guidelines, such as the Declaration of Helsinki, the Council for International Organizations of Medical Sciences (CIOMS) guidelines, and the International Council on Harmonisation Good Clinical Practice guidelines, as well as national and local laws where research is taking place. Partnering with local researchers and communities is essential not only to demonstrate respect but also to facilitate successful clinical research.

INTERNATIONAL CLINICAL EXPERIENCES

Many physicians and trainees gain valuable experience providing patient care in international settings through international training opportunities or volunteering for humanitarian or other international clinical work. Such arrangements, however, raise ethical challenges—for example, as a result of differences in beliefs about health and illness, expectations regarding health care and physicians' roles, standards of clinical practice, resource limitations, and norms for disclosure of serious diagnoses. Additional dilemmas arise if visiting physicians and trainees take on responsibilities beyond their expertise or if donated drugs and equipment are not appropriate to local needs. Visiting physicians and trainees should prepare well for these experiences, receive training and mentoring, learn about cultural and clinical practices in the host community, respect local customs and values, collaborate closely with local professionals and staff, and be explicit and humble about their own skills, knowledge, and limits. Leaders of global health field experiences should ensure that participating physicians receive training on ethical and cultural issues, as well as mentoring, backup, and debriefing upon return home.

CONCLUSION

Ethical issues are common in clinical medicine and occur in circumstances that may be foreseeable, novel, or unexpected. Physicians address these

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 11: Ethical Issues in Clinical Medicine, Christine Grady; Bernard Lo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FURTHER READING

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Ethical issues are common in clinical medicine and occur in circumstances that may be foreseeable, novel, or unexpected. Physicians address these ethical issues by being prepared, informed, and thoughtful and using appropriate available resources.

FURTHER READING

Beauchamp T, Childress J: *Principles of Biomedical Ethics*, 8th ed. New York, Oxford University Press, 2019.

Dejong C et al: An ethical framework for allocating scarce medications for COVID-19 in the US. JAMA 323:2367, 2020. [PubMed: 32412580]

Matheny M et al (eds): *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril.* NAM Special Publication. Washington, DC, National Academy of Medicine, 2019.

Ulrich C, Grady C: *Moral Distress in the Health Professions*. Cham, Switzerland, Springer-Nature International, 2018.

Wasserman J et al: Responding to unprofessional behavior by trainees: a "just culture" framework. N Engl J Med 382:773, 2020. [PubMed: 32074428]

Wicclair MR: Conscientious objection, moral integrity, and professional obligations. Perspect Biol Med 62:543, 2019. [PubMed: 31495797]

Downloaded 2022-1-3 5:41 P Your IP is 141.216.78.40
Chapter 11: Ethical Issues in Clinical Medicine, Christine Grady; Bernard Lo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 12: Palliative and End-of-Life Care

Ezekiel J. Emanuel

EPIDEMIOLOGY

CAUSES OF DEATH

In 2019, 2,854,838 individuals died in the United States **(Table 12-1)**. Approximately 74% of these deaths occurred in those aged ≥65 years. The epidemiology of death has changed significantly since 1900 and even since 1980. In 1900, heart disease caused ~8% of all deaths, and cancer accounted for <4% of all deaths. In 1980, heart disease accounted for 38.2% of all deaths, cancer 20.9%, and cerebrovascular disease 8.6% of all deaths. By 2019, there had been a dramatic drop in deaths from cardiovascular and cerebrovascular diseases. In 2019, 23.1% of all deaths were from cardiovascular disease and just 5.3% from cerebrovascular disease. Deaths attributable to cancer, however, had increased slightly to 21.0%. The proportions of deaths due to chronic lower respiratory disease, diabetes, Alzheimer's, and suicide have increased. Interestingly, in 2019, HIV/AIDS accounted for <0.18% of all U.S. deaths. While unlikely to continue being a leading cause of death in the future, COVID-19 was also the cause for >600,000 deaths in 2020–2021, and the official figure is almost certainly an undercount of the actual death toll.

TABLE 12-1

Ten Leading Causes of Death in the United States and Britain

CAUSE OF DEATH	UNITED STATES (2019)		ENGLAND AND WALES (2019)	
	NUMBER OF DEATHS, ALL AGES (%)	NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE	NUMBER OF DEATHS, ALL AGES (%)	NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE
All deaths	2,854,838	2,117,332	530,841	449,047
Heart diseasea	659,041 (23.1)	531,583 (25.1)	87,095 (16.4)	74,967 (16.7)
Malignant neoplasms	599,601 (21.0)	435,462 (20.6)	147,419 (27.8)	118,982 (26.5)
Chronic lower respiratory diseases	156,979 (5.5)	133,246 (6.3)	31,221 (5.9)	28,235 (6.3)
Accidents	173,040 (6.1)	60,527 (2.9)	15,141 (2.9)	8999 (2.0)
Cerebrovascular diseases	150,005 (5.3)	129,193 (6.1)	29,816 (5.6)	27,210 (6.0)
Alzheimer's disease	121,499 (4.3)	120,090 (5.7)	20,400 (3.8)	20,279 (4.5)
Diabetes mellitus	87,647 (3.1)	62,397 (2.9)	6528 (1.2)	5552 (1.2)
Influenza and pneumonia	49,783 (1.7)	40,399 (1.9)	26,398 (5.0)	24,269 (5.4)
Nephritis, nephritic	51,565 (1.8)	42,230 (2.0)	3575 (0.7)	3323 (0.7)

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 12-1

Ten Leading Causes of Death in the United States and Britain

CAUSE OF DEATH	UNITED STATES (2019)		ENGLAND AND WALES (2019)	
	NUMBER OF DEATHS, ALL AGES (%)	NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE	NUMBER OF DEATHS, ALL AGES (%)	NUMBER OF DEATHS, PEOPLE ≥65 YEARS OF AGE
All deaths	2,854,838	2,117,332	530,841	449,047
Heart diseasea	659,041 (23.1)	531,583 (25.1)	87,095 (16.4)	74,967 (16.7)
Malignant neoplasms	599,601 (21.0)	435,462 (20.6)	147,419 (27.8)	118,982 (26.5)
Chronic lower respiratory diseases	156,979 (5.5)	133,246 (6.3)	31,221 (5.9)	28,235 (6.3)
Accidents	173,040 (6.1)	60,527 (2.9)	15,141 (2.9)	8999 (2.0)
Cerebrovascular diseases	150,005 (5.3)	129,193 (6.1)	29,816 (5.6)	27,210 (6.0)
Alzheimer's disease	121,499 (4.3)	120,090 (5.7)	20,400 (3.8)	20,279 (4.5)
Diabetes mellitus	87,647 (3.1)	62,397 (2.9)	6528 (1.2)	5552 (1.2)
Influenza and pneumonia	49,783 (1.7)	40,399 (1.9)	26,398 (5.0)	24,269 (5.4)
Nephritis, nephritic syndrome, nephrosis	51,565 (1.8)	42,230 (2.0)	3575 (0.7)	3323 (0.7)
Intentional self-harm	47,511 (1.7)	—	4832 (0.9)	751 (0.2)

[a]Calculated using International Classification of Diseases codes I00–I09, I11, I13, I20–I51.

Source: National Center for Health Statistics (United States, 2019), *http://www.cdc.gov/nchs;* National Statistics (Great Britain, 2019), *http://www.statistics.gov.uk*.

This change in the epidemiology of death is also reflected in the costs of illness. In the United States, ~84% of all health care spending goes to patients with chronic illnesses, and 12% of total personal health care spending—slightly less than $400 billion in 2015—goes to the 0.83% of the population in the last year of their lives.

In upper-middle- and upper-income countries, an estimated 70% of all deaths are preceded by a disease or condition, making it reasonable to plan for dying in the foreseeable future. Cancer has served as the paradigm for terminal care, but it is not the only type of illness with a recognizable and predictable terminal phase. Since heart failure, chronic obstructive pulmonary disease (COPD), chronic liver failure, dementia, and many other conditions have recognizable terminal phases, a systematic approach to end-of-life care should be part of all medical specialties. Many patients with chronic illness–related symptoms and suffering also can benefit from palliative care regardless of prognosis. Ideally, palliative care should be considered part of comprehensive care for all chronically ill patients. Strong evidence demonstrates that palliative care can be improved by coordination between caregivers, doctors, and patients for advance care planning, as well as dedicated teams of physicians, nurses, and other providers.

SITE OF DEATH

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Most people have a strong preference to die at home, and in many countries, including the Netherlands, Sweden, and other countries like the United States, most of the hospital, as patients and their families use their own homes as the preferred site of death. In the early 1980s, ~70% of American cancer patients died in the hospital. Today, that percentage is ~25% (Fig. 12-1). A recent report shows that since 2000, there has been a shift in the United States from inpatient to home deaths, especially for patients

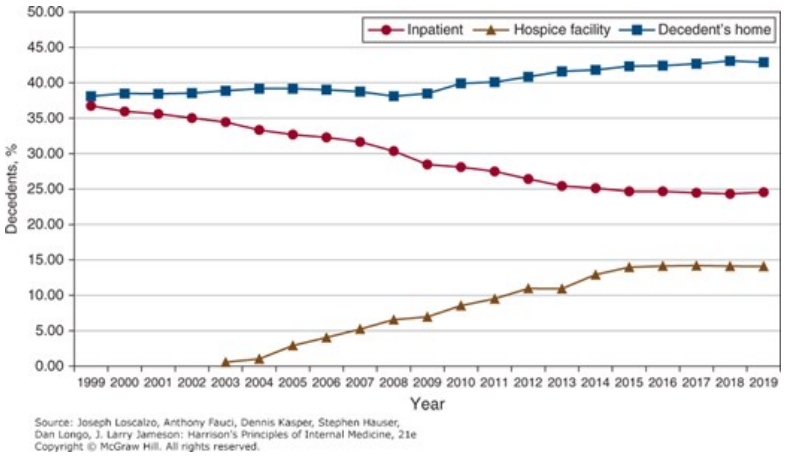
chronic illness–related symptoms and suffering also can benefit from palliative care regardless of prognosis. Ideally, palliative care should be

SITE OF DEATH

Where patients die varies by country. In Belgium and Canada, for instance, over half of all cancer patients still die in the hospital. The past few decades have seen a steady shift, both in the United States and other countries like the Netherlands, out of the hospital, as patients and their families list their own homes as the preferred site of death. In the early 1980s, ~70% of American cancer patients died in the hospital. Today, that percentage is ~25% (Fig. 12-1). A recent report shows that since 2000, there has been a shift in the United States from inpatient to home deaths, especially for patients with cancer, COPD, and dementia. For instance, among Medicare beneficiaries, 30.1% of deaths due to cancer in 2000 occurred in acute care hospitals; by 2009, this figure had dropped to 22.1%; by 2015, it was 19.8%.

FIGURE 12-1

Graph showing trends in cancer decedents' site of death 1999–2019. (*Source: Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2019 on CDC WONDER Online Database. http://wonder.cdc.gov.*)



Paradoxically, while deaths in acute care hospitals have declined in the United States since 2000, both hospitalizations in the last 90 days of life and—even more troublingly—admission to the intensive care unit (ICU) in the last 30 days have actually increased. Over 40% of cancer patients in the United States are admitted to the ICU in their last 6 months of life, and >25% of cancer patients are admitted to the hospital in the last 30 days.

The shift in deaths out of the hospital has been accompanied by an increase in the use of hospice in the United States. In 2000, 21.6% of Medicare decedents used hospice at the time of death; by 2009, 42.2% were using hospice; and by 2018, 50.7% of Medicare decedents were enrolled in hospice at the time of death. Among cancer patients, ~60% were using hospice at the time of death. Hospice is also increasingly being used by noncancer patients. Today, cancer patients constitute ~20% of hospice users. But since 2014, the proportion of patients with other diagnoses using hospice has grown substantially, including those with circulatory/heart disease (17.4% in 2018 vs 13.8% in 2014), stroke (9.5% vs 6.2%), and respiratory disease (11.0% vs 9.4%). Of 2018 Medicare hospice decedents, 51.5% died at home, 17.4% in a nursing facility, 12.8% in a hospice inpatient facility, and 12.3% in assisted living.

Unfortunately, significant racial disparities exist in end-of-life care and the use of hospice, especially for noncancer deaths. Racial and ethnic minorities are less likely to receive hospice services than white decedents and are more likely to receive invasive or aggressive care in end-of-life treatment. Of people who died of head and neck cancers between 1999 and 2017, African Americans and Asians/Pacific Islanders were less likely to die at home or in hospice. Among Medicare beneficiaries who had a pancreatectomy for pancreatic cancer and lived at least 30 days, racial and ethnic minority patients remained 22% less likely than white patients to initiate hospice before death.

In 2008, for the first time, the American Board of Medical Specialties (ABMS) offered certification in hospice and palliative medicine. With the shortening of hospital stays, many serious conditions are now being treated at home or on an outpatient basis. Consequently, providing optimal palliative and end-of-life care requires ensuring that appropriate services are available in a variety of settings, including noninstitutional settings.

HOSPICE AND THE PALLIATIVE CARE FRAMEWORK

Central to this type of care is an interdisciplinary team approach that typically encompasses pain and symptom management, spiritual and

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

switching to hospice in the patient's last few months. This approach avoids leaving hospice until the very end by introducing palliative care earlier, thereby allowing patients and families time to accommodate and transition. Phasing palliative care into end-of-life care means that patients will often

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Central to this type of care is an interdisciplinary team approach that typically encompasses pain and symptom management, spiritual and psychological care for the patient, and support for family caregivers during the patient's illness and the bereavement period.

One of the more important changes in this field is beginning palliative care many months before death in order to focus on symptom relief and then switching to hospice in the patient's last few months. This approach avoids leaving hospice until the very end by introducing palliative care earlier, thereby allowing patients and families time to accommodate and transition. Phasing palliative care into end-of-life care means that patients will often receive palliative interventions long before they are formally diagnosed as terminally ill, or likely to die within 6 months.

Fundamental to ensuring quality palliative and end-of-life care is a focus on four broad domains: (1) physical symptoms; (2) psychological symptoms; (3) social needs that include interpersonal relationships, caregiving, and economic concerns; and (4) existential or spiritual needs.

ASSESSMENT AND CARE PLANNING

Comprehensive Assessment

Standardized methods for conducting a comprehensive assessment focus on evaluating the patient's condition in all four domains affected by the illness: physical, psychological, social, and spiritual.

A comprehensive assessment should follow a modified version of the traditional medical history and physical examination and should emphasize both physical and mental symptoms. Questions should aim to elucidate symptoms, discern sources of suffering, and gauge how much those symptoms interfere with the patient's quality of life. Standardized and repeated assessments to evaluate the effectiveness of interventions are critical. Thus, clinicians should use shorter, validated instruments, such as (1) the revised Edmonton Symptom Assessment Scale; (2) Condensed Memorial Symptom Assessment Scale (MSAS); (3) MD Anderson Brief Symptom Inventory; (4) Rotterdam Symptom Checklist; (5) Symptom Distress Scale; (6) Patient-Reported Outcomes Measurement Information System; and (7) Interactive Symptom Assessment and Collection (ISAAC) tool.

MENTAL HEALTH

With respect to mental health, many practices use the Patient Health Questionnaire-9 (PHQ-9) to screen for depression and the Generalized Anxiety Disorder-7 (GAD-7) to screen for anxiety. Using such tools ensures that the assessment is comprehensive and does not focus excessively on only pain.

INVASIVE TESTS

Invasive tests are best avoided in end-of-life care, and even minimally invasive tests should be evaluated carefully for their benefit-to-burden ratio for the patient. Aspects of the physical examination that are uncomfortable and unlikely to yield useful information that change patient management should be omitted.

SOCIAL NEEDS

Health care providers should also assess the status of important relationships, financial burdens, caregiving needs, and access to medical care. Relevant questions include the following: *How often is there someone to feel close to? How has this illness been for your family? How has it affected your relationships? How much help do you need with things like getting meals and getting around? How much trouble do you have getting the medical care you need?*

EXISTENTIAL NEEDS

To determine a patient's existential needs, providers should assess distress, the patient's sense of emotional and existential well-being, and whether the patient believes he or she has found purpose or meaning. Helpful assessment questions can include the following: *How much are you able to find meaning since your illness began? What things are most important to you at this stage?*

PERCEPTION OF CARE

In addition, it can be helpful to ask how the patient perceives his or her care: *How much do you feel your doctors and nurses respect you? How clear is the information from us about what to expect regarding your illness? How much do you feel that the medical care you are getting fits with your goals?* If concern is detected in any of these areas, deeper evaluative questions are warranted.

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

collectively called "bad news" situations—in which empathic and effective communication skills are essential. Those moments include the sharing of a terminal diagnosis with the patient and/or family, the discussion of the patient's prognosis and any treatment failures, the consideration of

concern is detected in any of these areas, deeper evaluative questions are warranted.

Communication

Particularly when an illness is life-threatening, there exists the potential for many emotionally charged and potentially conflict-creating moments—collectively called "bad news" situations—in which empathic and effective communication skills are essential. Those moments include the sharing of a terminal diagnosis with the patient and/or family, the discussion of the patient's prognosis and any treatment failures, the consideration of deemphasizing efforts to cure and prolong life while focusing more on symptom management and palliation, advance care planning, and the patient's actual death. Although these conversations can be difficult, research indicates that end-of-life discussions can lead to earlier hospice referrals, rather than overly aggressive treatment, ultimately benefiting quality of life for patients and improving the bereavement process for families.

Just as surgeons prepare for major operations and investigators rehearse a presentation of research results, physicians and health care providers caring for patients with significant or advanced illnesses should develop a standardized approach for sharing important information and planning interventions. In addition, physicians must be aware that families often care not only about how prepared the physician was to deliver bad news, but also the setting in which it was delivered. For instance, one study found that 27% of families making critical decisions for patients in an ICU desired better and more private physical space to communicate with physicians.

One structured seven-step procedure for communicating bad news goes by the acronym P-SPIKES: (1) ***p***repare for the discussion, (2) ***s***et up a suitable environment, (3) begin the discussion by finding out what the ***p***atient and/or family understand, (4) determine how they will comprehend new ***i***nformation best and how much they want to know, (5) provide needed new ***k***nowledge accordingly, (6) allow for ***e***motional responses, and (7) ***s***hare plans for the next steps in care **(Table 12-2)**.

TABLE 12-2

Elements of Communicating Bad News—The P-SPIKES Approach

ACRONYM	STEPS	AIM OF THE INTERACTION	PREPARATIONS, QUESTIONS, OR PHRASES
P	Preparation	Mentally prepare for the interaction with the patient and/or family.	Review what information needs to be communicated. Plan how you will provide emotional support. Rehearse key steps and phrases in the interaction.
S	Setting of the interaction	Ensure the appropriate setting for a serious and potentially emotionally charged discussion.	Ensure that patient, family, and appropriate social supports are present. Devote sufficient time. Ensure privacy and prevent interruptions by people or beeper. Bring a box of tissues.
P	Patient's perception and preparation	Begin the discussion by establishing the baseline and whether the patient and family can grasp the information. Ease tension by having the patient and family contribute.	Start with open-ended questions to encourage participation. Possible questions to use: *What do you understand about your illness?* *When you first had symptom X, what did you think it might be?* *What did Dr. X tell you when he or she sent you here?* *What do you think is going to happen?*
I	Invitation and information needs	Discover what information needs the patient and/or family have and what limits they want regarding the bad information.	Possible questions to use: *If this condition turns out to be something serious, do you want to know?* *Would you like me to tell you all the details of your condition? If not, who would you like me to talk to?*
K	Knowledge of the condition	Provide the bad news or other information to the patient and/or family sensitively.	Do not just dump the information on the patient and family. Check for patient and family understanding. Possible phrases to use: *I feel badly to have to tell you this, but*
E	Empathy and	Identify the cause of the emotions—e.g., poor prognosis	Strong feelings in reaction to bad news are normal. Acknowledge what the patient and family are feeling

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

ACRONYM	STEPS	AIM OF THE INTERACTION	PREPARATIONS, QUESTIONS, OR PHRASES
P	Preparation	Mentally prepare for the interaction with the patient and/or family.	Review what information needs to be communicated. Plan how you will provide emotional support. Rehearse key steps and phrases in the interaction.
S	Setting of the interaction	Ensure the appropriate setting for a serious and potentially emotionally charged discussion.	Ensure that patient, family, and appropriate social supports are present. Devote sufficient time. Ensure privacy and prevent interruptions by people or beeper. Bring a box of tissues.
P	Patient's perception and preparation	Begin the discussion by establishing the baseline and whether the patient and family can grasp the information. Ease tension by having the patient and family contribute.	Start with open-ended questions to encourage participation. Possible questions to use: *What do you understand about your illness?* *When you first had symptom X, what did you think it might be?* *What did Dr. X tell you when he or she sent you here?* *What do you think is going to happen?*
I	Invitation and information needs	Discover what information needs the patient and/or family have and what limits they want regarding the bad information.	Possible questions to use: *If this condition turns out to be something serious, do you want to know?* *Would you like me to tell you all the details of your condition? If not, who would you like me to talk to?*
K	Knowledge of the condition	Provide the bad news or other information to the patient and/or family sensitively.	Do not just dump the information on the patient and family. Check for patient and family understanding. Possible phrases to use: *I feel badly to have to tell you this, but…* *Unfortunately, the tests showed…* *I'm afraid the news is not good…*
E	Empathy and exploration	Identify the cause of the emotions—e.g., poor prognosis. Empathize with the patient's and/or family's feelings. Explore by asking open-ended questions.	Strong feelings in reaction to bad news are normal. Acknowledge what the patient and family are feeling. Remind them such feelings are normal, even if frightening. Give them time to respond. Remind the patient and family you won't abandon them. Possible phrases to use: *I imagine this is very hard for you to hear.* *You look very upset. Tell me how you are feeling.* *I wish the news were different.* *We'll do whatever we can to help you.*
S	Summary and planning	Delineate for the patient and the family the next steps, including additional tests or interventions.	It is the unknown and uncertain that can increase anxiety. Recommend a schedule with goals and landmarks. Provide your rationale for the patient and/or family to accept (or reject). If the patient and/or family are not ready to discuss the next steps, schedule a follow-up visit.

Source: Adapted from R Buckman: *How to Break Bad News: A Guide for Health Care Professionals.* Baltimore, Johns Hopkins University Press, 1992.

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Major barriers to providing high-quality palliative and end-of-life care include the difficulty in determining an accurate prognosis and the emotional resistance of patients and their families to accepting the implications of a poor prognosis. A practical solution to these barriers is to integrate palliative

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Source: Adapted from R Buckman: *How to Break Bad News: A Guide for Health Care Professionals.* Baltimore, Johns Hopkins University Press, 1992.

Continuous Goal Assessment

Major barriers to providing high-quality palliative and end-of-life care include the difficulty in determining an accurate prognosis and the emotional resistance of patients and their families to accepting the implications of a poor prognosis. A practical solution to these barriers is to integrate palliative care interventions or home visits from a palliative care visiting nurse months before the estimated final 6 months of life. Under this approach, palliative care no longer conveys the message of failure, having no more treatments, or "giving up hope." The transition from palliative to end-of-life care or hospice also feels less hasty and unexpected to the family. Fundamental to integrating palliative care with curative therapy is the inclusion of a continuous goal assessment as part of the routine patient reassessments that occur at most patient-physician encounters.

Goals for care are numerous, ranging from curing a specific disease, to prolonging life, to relieving a particular symptom, to adapting to a progressive disability without disrupting the family, to finding peace of mind or personal meaning, to dying in a manner that leaves loved ones with positive memories. Discerning a patient's goals for care can be approached through a seven-step protocol: (1) ensure that medical and other information is as complete as reasonably possible and is understood by all relevant parties (see above); (2) explore what the patient and/or family is hoping for, while also identifying relevant and realistic goals; (3) share all the options with the patient and family; (4) respond with empathy as they adjust to changing expectations; (5) make a plan that emphasizes what can be done to achieve the realistic goals; (6) follow through with the plan; and (7) periodically review the plan and consider at every encounter whether the goals of care should be revised with the patient and/or family. Each of these steps need not be followed in rote order, but together they provide a helpful framework for interactions with patients and their families regarding their goals for care. Such interactions can be especially challenging if a patient or family member has difficulty letting go of an unrealistic goal. In such cases, the provider should help them refocus on more realistic goals and should also suggest that while it is fine to hope for the best, it is still prudent to plan for other outcomes as well.

Advance Care Planning

PRACTICES

Advance care planning is the process of planning for future medical care in case the patient becomes incapable of making medical decisions. A 2010 study of adults aged ≥60 who died between 2000 and 2006 found that while 42% of adults were required to make treatment decisions in their final days of life, 70% lacked decision-making capacity. Among those lacking decision-making capacity, approximately one-third did not have advance planning directives. Ideally, such planning would occur before a health care crisis or the terminal phase of an illness. Unfortunately, diverse barriers prevent this. Approximately 80% of Americans endorse advance care planning and living wills. However, according to a 2013 Pew survey, only 35% of adults have written down their end-of-life wishes. Other studies report that even fewer Americans—with some estimates as low as 26% of adults—have filled out advance care directives. A review of studies suggests that the percentage of Americans who had written advance directives did not change between 2011 and 2016 and remains slightly over one-third of Americans. Larger numbers of adults, between 50 and 70%, claim to have talked with someone about their treatment wishes. Americans aged 65 and older are more likely to complete an advance directive compared to younger adults (46% vs 32%).

Effective advance care planning should follow six key steps: (1) introducing the topic, (2) structuring a discussion, (3) reviewing plans that have been discussed by the patient and family, (4) documenting the plans, (5) updating them periodically, and (6) implementing the advance care directives **(Table 12-3)**. Two of the main barriers to advance care planning are problems in raising the topic and difficulty in structuring a succinct discussion. Raising the topic can be done efficiently as a routine matter, noting that it is recommended for all patients, analogous to purchasing insurance or estate planning. Many of the most difficult cases have involved unexpected, acute episodes of brain damage in young individuals.

TABLE 12-3

Steps in Advance Care Planning

STEP	GOALS TO BE ACHIEVED AND MEASURES TO COVER	USEFUL PHRASES OR POINTS TO MAKE
Introduce advance care	Ask the patient what he or she knows about advance care	*I'd like to talk with you about something I try to discuss with all my patients. It's called advance care planning. In fact, I feel that this is such an important topic that I have done this myself. Are you*
	Indicate that you as a physician	*Have you thought about the type of care you would want if you ever became too sick to speak for*

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

advance care planning	knows about advance care planning and if he or she has already completed an advance care directive.	*care planning. In fact, I feel that this is such an important topic that I have done this myself. Are you familiar with advance care planning or living wills?*
	Indicate that you as a physician have completed advance care planning.	*Have you thought about the type of care you would want if you ever became too sick to speak for yourself? That is the purpose of advance care planning.*
	Indicate that you try to perform advance care planning with all patients regardless of prognosis.	*There is no change in health that we have not discussed. I am bringing this up now because it is sensible for everyone, no matter how well or ill, old or young.*
	Explain the goals of the process as empowering the patient and ensuring that you and the proxy understand the patient's preferences.	*Have many copies of advance care directives available, including in the waiting room, for patients and families.*
	Provide the patient relevant literature, including the advance care directive that you prefer to use.	*Know resources for state-specific forms (available at www.nhpco.org).*
	Recommend the patient identify a proxy decision-maker who should attend the next meeting.	
Have a structured discussion of scenarios with the patient	Affirm that the goal of the process is to follow the patient's wishes if the patient loses decision-making capacity.	*Use a structured worksheet with typical scenarios.*
	Elicit the patient's overall goals related to health care. Elicit the patient's preferences for specific interventions in a few salient and common scenarios. Help the patient define the threshold for withdrawing and withholding interventions. Define the patient's preference for the role of the proxy.	*Begin the discussion with persistent vegetative state and consider other scenarios, such as recovery from an acute event with serious disability; then ask the patient about his or her preferences regarding specific interventions, such as ventilators, artificial nutrition, and CPR; finally, proceeding to less invasive interventions, such as blood transfusions and antibiotics.*
Review the patient's preferences	After the patient has made choices of interventions, review them to ensure they are consistent and the proxy is aware of them.	
Document the patient's	Formally complete the advance care directive and have a witness	
	the proxy.	

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

the patient's preferences	care directive and have a witness sign it.	
	Provide a copy for the patient and the proxy.	
	Insert a copy into the patient's medical record and summarize it in a progress note.	
Update the directive	Periodically, and with major changes in health status, review the directive with the patient and make any modifications.	
Apply the directive	The directive goes into effect only when the patient becomes unable to make medical decisions for himself or herself.	
	Reread the directive to be sure about its content.	
	Discuss your proposed actions based on the directive with the proxy.	

Abbreviation: CPR, cardiopulmonary resuscitation.

Structuring a focused discussion is an important communication skill. To do so, a provider must first identify the health care proxy and recommend his or her involvement in the advance care planning process. Next, a worksheet must be selected that has been demonstrated to produce reliable and valid expressions of patient preferences, and the patient and proxy must be oriented to it. Such worksheets exist for both general and disease-specific situations. The provider should then discuss with the patient and proxy one example scenario to demonstrate how to think about the issues. It is often helpful to begin with a scenario in which the patient is likely to have settled preferences for care, such as being in a persistent vegetative state. Once the patient's preferences for interventions in this scenario are determined, the provider should suggest that the patient and proxy discuss and complete the worksheet for each other. If appropriate, the patient and proxy should consider involving other family members in the discussion. During a subsequent return visit, the provider should go over the patient's preferences, checking and resolving any inconsistencies. After having the patient and proxy sign the document, the provider should place the document in the patient's medical chart and make sure that copies are provided to relevant family members and care sites. Since patients' preferences can change, these documents must be reviewed periodically.

TYPES OF DOCUMENTS

Advance care planning documents are of two broad types. The first includes living wills, also known as instructional directives; these are advisory documents that describe the types of decisions that should direct a patient's care. Some are more specific, delineating different scenarios and interventions for the patient to choose from. Among these, some are for general use and others are designed for use by patients with a specific type of disease, such as cancer, renal failure, or HIV. Less specific directives can be general statements, such as not wanting life-sustaining interventions, or forms that describe the values that should guide specific discussions about terminal care. The second type of advance directive allows the designation of a health care proxy (sometimes also referred to as a durable attorney for health care), an individual selected by the patient to make decisions. The choice is not either/or; a combined directive that includes a living will and designates a proxy is often used, and the directive should indicate clearly whether the specified patient preferences or the proxy's choice takes precedence if they conflict. Some states have begun to put into practice a "Physician Orders for Life-Sustaining Treatment (POLST)" directive, which builds on communication between providers and patients by including

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

A potentially misleading distinction relates to statutory, as opposed to advisory, documents. Statutory documents are drafted to fulfill relevant state laws. Advisory documents are drafted to reflect the patient's wishes. Both are legal, the former under state law and the latter under common or

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Whether the specified patient preferences of the proxy's choice takes precedence if they conflict. Some states have begun to put into practice a "Physician Orders for Life-Sustaining Treatment (POLST)" directive, which builds on communication between providers and patients by including guidance for end-of-life care in a color-coordinated form that follows the patient across treatment settings. The procedures for completing advance care planning documents vary according to state law.

A potentially misleading distinction relates to statutory, as opposed to advisory, documents. Statutory documents are drafted to fulfill relevant state laws. Advisory documents are drafted to reflect the patient's wishes. Both are legal, the former under state law and the latter under common or constitutional law.

LEGAL ASPECTS

As of 2021, 48 states and the District of Columbia had enacted living will legislation. Massachusetts and Michigan are the two states without living will legislation. Indiana has a life-prolonging procedures declaration. States differ in the requirements for advanced directives, including whether they need to be witnessed and, if so, by how many witnesses and whether they need to be notarized. Importantly, in 25 states, the laws state that the living will is not valid if a woman is pregnant. All states except Alaska have enacted durable power of attorney for health care laws that permit patients to designate a proxy decision-maker with authority to terminate life-sustaining treatments. Only in Alaska does the law prohibit proxies from terminating life-sustaining treatments for pregnant women.

The U.S. Supreme Court has ruled that patients have a constitutional right to decide any issues related to refusing or terminating medical interventions, including life-sustaining interventions, and that mentally incompetent patients can exercise this right by providing "clear and convincing evidence" of their preferences. Since advance care directives permit patients to provide such evidence, commentators agree that they are constitutionally protected. Most commentators believe that a state is required to honor any clear advance care directive, regardless of whether it is written on an "official" form. Many states have enacted laws for the explicit purpose of honoring out-of-state directives. If a patient is not using a statutory form, it may be advisable to attach a statutory form to the advance care directive being used. State-specific forms are readily available free of charge for health care providers, patients, and families through the website of the National Hospice and Palliative Care Organization (*http://www.nhpco.org*).

REIMBURSEMENT

As of January 1, 2016, the Centers for Medicare and Medicaid Services amended the physician fee schedule to reimburse discussions of advance care planning under Current Procedural Terminology codes 99497 and 99498. The session must be voluntary and include an explanation of advance care planning but need not include a completed advance care document. There can be multiple bills for the discussion if it extends over several encounters. A study found that patients who engaged in a billed advance care planning encounter were more likely to be enrolled in hospice and less likely to receive intensive therapies, despite being more likely to be hospitalized in the ICU. However, a billing incentive in and of itself may not increase advance care planning discussions by clinicians. In 2016, just 1.6% of Medicare Advantage patients had a discussion of advance care planning that was billed. Factors beyond reimbursement, such as clinicians' lack of comfort and skill in carrying out advance care planning discussions and lack of time, appear to impede discussions of advance care planning.

INTERVENTIONS

PHYSICAL SYMPTOMS AND THEIR MANAGEMENT

Great emphasis has been placed on addressing dying patients' pain. In order to emphasize its importance, pain assessment has frequently been included as the fifth vital sign. Heightened consideration of pain has been advocated by large health care systems such as the Veterans' Administration and accrediting bodies such as The Joint Commission. Although this embrace of pain has been symbolically important, available data suggest that making pain the fifth vital sign does not lead to improved pain management practices. In light of the opioid crisis in the United States, the emphasis on pain management has begun to be reexamined. For instance, in 2017 draft standards, The Joint Commission recommends nonpharmacologic pain treatment as well as identification of psychosocial risk factors for addiction. Importantly, good palliative care requires much more than good pain management. The frequency of symptoms varies by disease and other factors. The most common physical and psychological symptoms among all terminally ill patients include pain, fatigue, insomnia, anorexia, dyspnea, depression, anxiety, nausea, and vomiting. In the last days of life, terminal delirium is also common. Assessments of patients with advanced cancer have shown that patients experienced an average of 11.5 different physical and psychological symptoms **(Table 12-4)**.

TABLE 12-4

PHYSICAL SYMPTOMS	PSYCHOLOGICAL SYMPTOMS

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 12-4

Common Physical and Psychological Symptoms of Terminally Ill Patients

PHYSICAL SYMPTOMS	PSYCHOLOGICAL SYMPTOMS
Pain	Anxiety
Fatigue and weakness	Depression
Dyspnea	Hopelessness
Insomnia	Meaninglessness
Dry mouth	Irritability
Anorexia	Impaired concentration
Nausea and vomiting	Confusion
Constipation	Delirium
Cough	Loss of libido
Swelling of arms or legs	
Itching	
Diarrhea	
Dysphagia	
Dizziness	
Fecal and urinary incontinence	
Numbness/tingling in hands/feet	

In the vast majority of cases, evaluations to determine the etiology of these symptoms should be limited to the history and physical examination. In some cases, radiologic or other diagnostic examinations will provide sufficient benefit in directing optimal palliative care to warrant the risks, potential discomfort, and inconvenience, especially to a seriously ill patient. Only a few of the common symptoms that present difficult management issues will be addressed in this chapter. **Additional information on the management of other symptoms, such as nausea and vomiting, insomnia, and diarrhea, can be found in Chaps. 45, 31, and 46, respectively. Information on the management of patients with cancer is provided in Chap. 69.**

Pain

FREQUENCY

The frequency of pain among terminally ill patients varies significantly. Cancer (~85%), congestive heart failure (CHF; ~75%), and AIDS have been associated with a higher prevalence of pain compared to other advanced illnesses, such as COPD (~45%), chronic kidney disease (~40%), and dementia (~40%). One meta-analysis of adults with advanced or terminal illness found pain prevalence of 30–94% in patients with cancer, compared to 21–77%

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

There are two types of pain: nociceptive and neuropathic. Nociceptive pain is further divided into somatic or visceral pain. *Somatic pain* is the result of

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

associated with a higher prevalence of pain compared to other advanced illnesses, such as COPD (~45%), chronic kidney disease (~40%), and dementia (~40%). One meta-analysis of adults with advanced or terminal illness found pain prevalence of 30–94% in patients with cancer, compared to 21–77% for COPD, 14–78% for CHF, 11–83% for end-stage renal disease, 14–63% for dementia, and 30–98% for AIDS.

ETIOLOGY

There are two types of pain: nociceptive and neuropathic. Nociceptive pain is further divided into somatic or visceral pain. *Somatic pain* is the result of direct mechanical or chemical stimulation of nociceptors and normal neural signaling to the brain. It tends to be localized, aching, throbbing, and cramping. The classic example is bone metastases. *Visceral pain* is caused by nociceptors in gastrointestinal (GI), respiratory, and other organ systems. It is a deep or colicky type of pain classically associated with pancreatitis, myocardial infarction, or tumor invasion of viscera. *Neuropathic pain* arises from disordered nerve signals. It is described by patients as burning, electrical, or shock-like pain. Classic examples are post-stroke pain, tumor invasion of the brachial plexus, and herpetic neuralgia.

ASSESSMENT

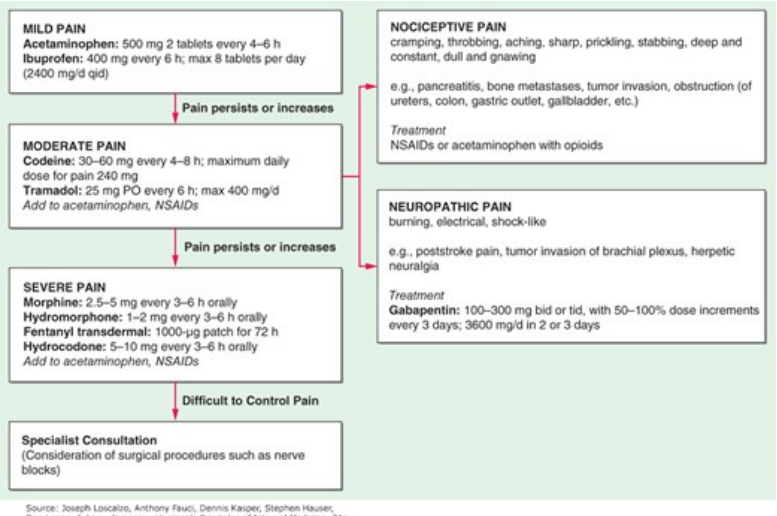
Pain is a subjective experience. Depending on the patient's circumstances, perspective, and physiologic condition, the same physical lesion or disease state can produce different levels of reported pain and need for pain relief. Systematic assessment includes eliciting the following: (1) type: throbbing, cramping, burning, etc.; (2) periodicity: continuous, with or without exacerbations, or incident; (3) location; (4) intensity; (5) modifying factors; (6) effects of treatments; (7) functional impact; and (8) impact on patient. Several validated pain assessment measures may be used, including the Visual Analogue Scale (VAS), the Brief Pain Inventory (BPI), or the Numerical Pain Rating Scale (NRS-11). Other scales have been developed for neuropathic pain, such as the Neuropathic Pain Scale and the DN4 Questionnaire. Frequent reassessments on a consistent scale are essential to assess the impact of and need to readjust interventions.

INTERVENTIONS

Interventions for pain must be tailored to each individual, with the goal of preempting chronic pain and relieving breakthrough pain. At the end of life, there is rarely reason to doubt a patient's report of pain. With the opioid crisis in the United States, there is more emphasis on making opioids one component of multimodal analgesia. Nevertheless, at the end of life, pain medications, especially opioids, remain the cornerstone of management **(Fig. 12-2)**. If they are failing and nonpharmacologic interventions—including radiotherapy and anesthetic or neurosurgical procedures such as peripheral nerve blocks or epidural medications—are required, a pain consultation is appropriate.

FIGURE 12-2

Terminal pain management flow chart. NSAIDs, nonsteroidal anti-inflammatory drugs.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Pharmacologic interventions still largely follow the World Health Organization three-step, "analgesic ladder" approach, which involves nonopioid analgesics, "mild" opioids, and "strong" opioids, with or without adjuvants **(Chap. 13)**. Nonopioid analgesics, especially nonsteroidal anti-inflammatory drugs (NSAIDs), are the initial treatments for mild pain. They work primarily by inhibiting peripheral prostaglandins and reducing inflammation but may also have central nervous system (CNS) effects. Additionally, NSAIDs have a ceiling effect. Ibuprofen, up to 2400 mg/d qid, has a proton pump inhibitor, should be used. Acetaminophen is an alternative in patients with a history of GI bleeding and can be used safely at up to 4 g/d qid. In patients with liver dysfunction due to metastases or other causes and in patients with heavy alcohol use, doses should be reduced.

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

inflammatory drugs (NSAIDs), are the initial treatments for mild pain. They work primarily by inhibiting peripheral prostaglandins and reducing inflammation but may also have central nervous system (CNS) effects. Additionally, NSAIDs have a ceiling effect. Ibuprofen, up to 2400 mg/d qid, has a minimal risk of causing bleeding and renal impairment and is a good initial choice. In patients with a history of severe GI or other bleeding, however, ibuprofen should be avoided. In patients with a history of mild gastritis or gastroesophageal reflux disease (GERD), acid-lowering therapy, such as a proton pump inhibitor, should be used. Acetaminophen is an alternative in patients with a history of GI bleeding and can be used safely at up to 4 g/d qid. In patients with liver dysfunction due to metastases or other causes and in patients with heavy alcohol use, doses should be reduced.

If nonopioid analgesics are insufficient, opioids should be introduced. Opioids primarily work by interacting with μ opioid receptors to activate pain-inhibitory neurons in the CNS, although they also interact variably with δ and κ receptors. Receptor agonists, such as morphine, codeine, and fentanyl, produce analgesia by activating pain-inhibitory neurons in the CNS. Partial agonists, such as buprenorphine, have a ceiling effect for analgesia and a lower potential for abuse. They are useful for postacute pain but should not be used for chronic pain in end-of-life care. Pure antagonists, such as naloxone and methylnaltrexone, are used for reversal of opioid effects.

Traditionally, "weak" opioids such as codeine were used first. If they failed to relieve pain after dose escalation, "strong" opioids like morphine were used in doses of 5–10 mg every 4 h. However, this breakdown between "weak" and "strong" opioids is no longer commonly accepted, with smaller doses of "stronger" opioids frequently being preferred over similar or larger doses of "weaker" opioids, and different pain syndromes having different preferred therapies. Regardless, nonopioid analgesics should be combined with opioids, as they potentiate the effect of opioids.

Importantly, the goal is to prevent patients from experiencing pain. Consequently, for continuous pain, opioids should be administered on a regular, around-the-clock basis consistent with their duration of analgesia, and the next dose should occur before the effect of the previous dose wears off. They should not be provided only when the patient experiences pain. Patients should also be provided rescue medication, such as liquid morphine, for breakthrough pain, generally at 20% of the baseline dose. Patients should be informed that using the rescue medication does not obviate the need to take the next standard dose of pain medication. If the patient's pain remains uncontrolled after 24 h and recurs before the next dose, requiring the patient to utilize the rescue medication, the daily opioid dose can be increased by the total dose of rescue medications used by the patient, or by 50% of the standing opioid daily dose for moderate pain and 100% for severe pain.

It is inappropriate to start with extended-release preparations. Instead, an initial focus on using short-acting preparations to determine how much is required in the first 24–48 h will allow clinicians to determine opioid needs. Once pain relief is obtained using short-acting preparations, the switch should be made to extended-release preparations. Even with a stable extended-release preparation regimen, the patient may experience incident pain, such as during movement or dressing changes. Short-acting preparations should be taken before such predictable episodes. Although less common, patients may have "end-of-dose failure" with long-acting opioids, meaning that they develop pain after 8 h in the case of an every-12-h medication. In these cases, a trial of giving an every-12-h medication every 8 h is appropriate.

Due to differences in opioid receptors, cross-tolerance among opioids is incomplete, and patients may experience different side effects with different opioids. Therefore, if a patient is not experiencing pain relief or is experiencing too many side effects, a change to another opioid preparation is appropriate. When switching, one should begin with 50–75% of the published equianalgesic dose of the new opioid.

Unlike NSAIDs, opioids have no ceiling effect; therefore, there is no maximum dose, no matter how many milligrams the patient is receiving. The appropriate dose is the dose needed to achieve pain relief. This is an important point for clinicians to explain to patients and families. Addiction or excessive respiratory depression is extremely unlikely in the terminally ill; fear of these side effects should neither prevent escalating opioid medications when the patient is experiencing insufficient pain relief nor justify using opioid antagonists.

Opioid side effects should be anticipated and treated preemptively. Nearly all patients experience constipation that can be debilitating (see below). Failure to prevent constipation often results in noncompliance with opioid therapy. The preferred treatment is prevention. Cathartics (senna 2 tbsp qHS), stool softeners (docusate 100 mg PO qd), and/or laxatives (laxtulose 30 mL qd) are considered first-line treatment. For refractory cases, opioid antagonists or other therapies, such as lubiprostone, should be considered.

Methylnaltrexone is the best-studied opioid antagonist for use in refractory opioid-induced constipation. It reverses opioid-induced constipation by blocking peripheral opioid receptors, but not central receptors, for analgesia. In placebo-controlled trials, it has been shown to cause laxation within 24 h of administration. As with the use of opioids, about a third of patients using methylnaltrexone experience nausea and vomiting, but unlike with opioid usage, tolerance usually develops within a week. Therefore, when one is beginning opioids, an antiemetic such as metoclopramide or a serotonin antagonist is often prescribed prophylactically and stopped after 1 week. Olanzapine has also been shown to have antinausea properties and can be effective in countering delirium or anxiety, with the advantage of some weight gain.

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Seriously ill patients who require chronic pain relief rarely become addicted. Suspicion of addiction should not be a reason to withhold pain

and can be effective in countering delirium or anxiety, with the advantage of some weight gain.

Drowsiness, a common side effect of opioids, also usually abates within a week. For refractory or severe cases, pharmacologic therapy should be considered. The best-studied agents are the psychostimulants dextroamphetamine, methylphenidate, and modafinil, although evidence regarding their efficacy is weak. Modafinil has the advantage of once-a-day dosing compared to methyphenidate's twice daily dosing.

Seriously ill patients who require chronic pain relief rarely become addicted. Suspicion of addiction should not be a reason to withhold pain medications from terminally ill patients. Nonetheless, patients and families may withhold prescribed opioids for fear of addiction or dependence. Physicians and health care providers should reassure patients and families that the patient will not become addicted to opioids if they are used as prescribed for pain relief; this fear should not prevent the patient from taking the medications around the clock. However, diversion of drugs for use by other family members or illicit sale may occur. It may be necessary to advise the patient and caregiver about secure storage of opioids. Contract writing with the patient and family can help. If that fails, transfer to a safe facility may be necessary.

Tolerance describes the need to increase medication dosage for the same pain relief without a concurrent change in disease. In the case of patients with advanced disease, the need for increasing opioid dosage for pain relief usually is caused by disease progression rather than tolerance. Physical dependence is indicated by symptoms resulting from the abrupt withdrawal of opioids and should not be confused with addiction.

In recent years, the potential dangers of opioid drugs have become increasingly apparent. To help mitigate the risk of these powerful drugs, several strategies should be used to reduce the risk of aberrant drug use. To start, all patients should be assessed for their individual levels of risk. While there are multiple surveys available, including the Opioid Risk Tool, none have gained widespread use or validation. In general, however, it is important to screen for prior substance abuse and major psychiatric disorders.

For patients deemed to be high risk, a multidisciplinary effort should be pursued to reduce the risk of adverse consequences, such as addiction and diversion. Prescribing strategies include selecting opioids with longer durations of action and lower street values, such as methadone, and prescribing smaller quantities with more frequent follow-up. Monitoring options include periodic urine screening and referral to pain specialists. In some cases, it may also be reasonable to consider not offering short-acting opioids for breakthrough pain. In no situation, however, should adequate pain relief be withheld due to risk.

Adjuvant analgesic medications are nonopioids that potentiate the analgesic effects of opioids. They are especially important in the management of neuropathic pain. Gabapentin, an anticonvulsant initially studied in the setting of herpetic neuralgia, is now the first-line treatment for neuropathic pain resulting from a variety of causes. It is begun at 100–300 mg bid or tid, with 50–100% dose increments every 3 days. Usually 900–3600 mg/d in two or three doses is effective. The combination of gabapentin and nortriptyline may be more effective than gabapentin alone. Two potential side effects of gabapentin to be aware of are confusion and drowsiness, especially in the elderly. Other effective adjuvant medications include pregabalin, which has the same mechanism of action as gabapentin but is absorbed more efficiently from the GI tract. Lamotrigine is a novel agent whose mechanism of action is unknown but has been shown to be effective. It is recommended to begin at 25–50 mg/d, increasing to 100 mg/d. Carbamazepine, a first-generation agent, has been proven effective in randomized trials for neuropathic pain. Other potentially effective anticonvulsant adjuvants include topiramate (25–50 mg qd or bid, rising to 100–300 mg/d) and oxcarbazepine (75–300 mg bid, rising to 1200 mg bid).

Glucocorticoids, preferably dexamethasone given once a day, can be useful in reducing inflammation that causes pain, while also elevating mood, energy, and appetite. Its main side effects include confusion, sleep difficulties, and fluid retention. Glucocorticoids are especially effective for bone pain and abdominal pain from distention of the GI tract or liver. Other drugs, including clonidine and baclofen, can be effective in providing pain relief. These drugs are adjuvants and generally should be used in conjunction with—not instead of—opioids. Methadone, carefully dosed because of its unpredictable half-life in many patients, has activity at the *N*-methyl-D-aspartamate (NMDA) receptor and is useful for complex pain syndromes and neuropathic pain. It is generally reserved for cases in which first-line opioids (morphine, oxycodone, hydromorphone) are either ineffective or unavailable.

Radiation therapy can treat bone pain from single metastatic lesions. Bone pain from multiple metastases can be amenable to radiopharmaceuticals such as strontium-89 and samarium-153. Bisphosphonates, such as pamidronate (90 mg every 4 weeks) and calcitonin (200 IU intranasally once or twice a day), also provide relief from bone pain but have multiday onsets of action.

Constipation

FREQUENCY

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Although hypercalcemia and other factors can cause constipation, it is most frequently a predictable consequence of the use of opioids for pain and dyspnea relief and of the anticholinergic effects of tricyclic antidepressants, as well as due to the inactivity and poor diets common among seriously ill

twice a day), also provide relief from bone pain but have multiday onsets of action.

Constipation is reported in up to 70–100% of patients requiring palliative care.

ETIOLOGY

Although hypercalcemia and other factors can cause constipation, it is most frequently a predictable consequence of the use of opioids for pain and dyspnea relief and of the anticholinergic effects of tricyclic antidepressants, as well as due to the inactivity and poor diets common among seriously ill patients. If left untreated, constipation can cause substantial pain and vomiting and also is associated with confusion and delirium. Whenever opioids and other medications known to cause constipation are used, preemptive treatment for constipation should be instituted.

ASSESSMENT

Assessing constipation can be difficult because people describe it differently. Four commonly used assessment scales are the Bristol Stool Form Scale, the Constipation Assessment Scale, the Constipation Visual Analogue Scale, and the Eton Scale Risk Assessment for Constipation. The Bowel Function Index can be used to quantify opioid-induced constipation. The physician should establish the patient's previous bowel habits, as well as any changes in subjective and objective qualities such as bloating or decreased frequency. Abdominal and rectal examinations should be performed to exclude impaction or an acute abdomen. Radiographic assessments beyond a simple flat plate of the abdomen in cases in which obstruction is suspected are rarely necessary.

INTERVENTION

Any measure to address constipation during end-of-life care should include interventions to reestablish comfortable bowel habits and to relieve pain or discomfort. Although physical activity, adequate hydration, and dietary treatments with fiber can be helpful, each is limited in its effectiveness for most seriously ill patients, and fiber may exacerbate problems in the setting of dehydration or if impaired motility is the etiology. Fiber is contraindicated in the presence of opioid use. Stimulant and osmotic laxatives, stool softeners, fluids, and enemas are the mainstays of therapy **(Table 12-5)**. To prevent constipation from opioids and other medications, a combination of a laxative and a stool softener (such as senna and docusate) should be used. If after several days of treatment a bowel movement has not occurred, a rectal examination to remove impacted stool and place a suppository is necessary. For patients with impending bowel obstruction or gastric stasis, octreotide to reduce secretions can be helpful. For patients in whom the suspected mechanism is dysmotility, metoclopramide can be helpful.

TABLE 12-5

Medications for the Management of Constipation

INTERVENTION	DOSE	COMMENT
Stimulant laxatives		These agents directly stimulate peristalsis and may reduce colonic absorption of water.
Prune juice	120–240 mL/d	Work in 6–12 h.
Senna (Senokot)	2–8 tablets PO bid	
Bisacodyl	5–15 mg/d PO, PR	
Osmotic laxatives		These agents are not absorbed. They attract and retain water in the gastrointestinal tract.
Lactulose	15–30 mL PO q4–8h	Lactulose may cause flatulence and bloating.
Magnesium hydroxide (Milk of Magnesia)	15–30 mL/d PO	Lactulose works in 1 day, magnesium products in 6 h.
Stool softeners		These agents work by increasing water secretion and as detergents, increasing water penetration into the stool.

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 12-5

Medications for the Management of Constipation

INTERVENTION	DOSE	COMMENT
Stimulant laxatives		These agents directly stimulate peristalsis and may reduce colonic absorption of water.
Prune juice	120–240 mL/d	Work in 6–12 h.
Senna (Senokot)	2–8 tablets PO bid	
Bisacodyl	5–15 mg/d PO, PR	
Osmotic laxatives		These agents are not absorbed. They attract and retain water in the gastrointestinal tract.
Lactulose	15–30 mL PO q4–8h	Lactulose may cause flatulence and bloating.
Magnesium hydroxide (Milk of Magnesia)	15–30 mL/d PO	Lactulose works in 1 day, magnesium products in 6 h.
Magnesium citrate	125–250 mL/d PO	
Stool softeners		These agents work by increasing water secretion and as detergents, increasing water penetration into the stool.
Sodium docusate (Colace)	300–600 mg/d PO	Work in 1–3 days.
Calcium docusate	300–600 mg/d PO	
Suppositories and enemas		
Bisacodyl	10–15 PR qd	
Sodium phosphate enema	PR qd	Fixed dose, 4.5 oz, Fleet's.

Nausea

FREQUENCY

Up to 70% of patients with advanced cancer have nausea, defined as the subjective sensation of wanting to vomit.

ETIOLOGY

Nausea and vomiting are both caused by stimulation at one of four sites: the GI tract, the vestibular system, the chemoreceptor trigger zone (CTZ), and the cerebral cortex. Medical treatments for nausea are aimed at receptors at each of these sites: the GI tract contains mechanoreceptors

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

cortex is anticipatory nausea before a dose of chemotherapy or other noxious stimuli.

Up to 70% of patients with advanced cancer have nausea, defined as the subjective sensation of wanting to vomit.

ACCESS Medicine

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Nausea and vomiting are both caused by stimulation at one of four sites: the GI tract, the vestibular system, the chemoreceptor trigger zone (CTZ), and the cerebral cortex. Medical treatments for nausea are aimed at receptors at each of these sites: the GI tract contains mechanoreceptors, chemoreceptors, and 5-hydroxytryptamine type 3 (5-HT_3) receptors; the vestibular system probably contains histamine and acetylcholine receptors; and the CTZ contains chemoreceptors, dopamine type 2 receptors, and 5-HT3 receptors. An example of nausea that most likely is mediated by the cortex is anticipatory nausea before a dose of chemotherapy or other noxious stimuli.

Specific causes of nausea include metabolic changes (liver failure, uremia from renal failure, hypercalcemia), bowel obstruction, constipation, infection, GERD, vestibular disease, brain metastases, medications (including antibiotics, NSAIDs, proton pump inhibitors, opioids, and chemotherapy), and radiation therapy. Anxiety can also contribute to nausea.

INTERVENTION

Medical treatment of nausea is directed at the anatomic and receptor-mediated cause revealed by a careful history and physical examination. When no specific cause of nausea is identified, many advocate beginning treatment with metoclopramide; a serotonin type 3 (5-HT_3) receptor antagonist such as ondansetron, granisetron, palonosetron, dolasetron, tropisetron, or ramosetron; or a dopamine antagonist such as chlorpromazine, haloperidol, or prochlorperazine. When decreased motility is suspected, metoclopramide can be an effective treatment. When inflammation of the GI tract is suspected, glucocorticoids, such as dexamethasone, are an appropriate treatment. For nausea that follows chemotherapy and radiation therapy, one of the 5-HT_3 receptor antagonists or neurokinin-1 antagonists, such as aprepitant or fosaprepitant, is recommended. Clinicians should attempt prevention of postchemotherapy nausea, rather than simply providing treatment after the fact. Current clinical guidelines recommend tailoring the strength of treatments to the specific emetic risk posed by a specific chemotherapy drug. When a vestibular cause (such as “motion sickness” or labyrinthitis) is suspected, antihistamines, such as meclizine (whose primary side effect is drowsiness), or anticholinergics, such as scopolamine, can be effective. In anticipatory nausea, patients can benefit from nonpharmacologic interventions, such as biofeedback and hypnosis. The most common pharmacologic intervention for anticipatory nausea is a benzodiazepine, such as lorazepam. As with antihistamines, drowsiness and confusion are the main side effects.

The use of medical marijuana or oral cannabinoids for palliative treatment of nausea is controversial, as there are no controlled trials showing its effectiveness for patients at the end of life. A 2015 meta-analysis showed “low-quality evidence suggesting that cannabinoids were associated with improvements in nausea and vomiting due to chemotherapy,” and such treatments are not as good as 5-HT_3 receptor antagonists and can sometimes even cause cannabis hyperemesis syndrome. Older patients, who compose the vast majority of dying patients, seem to tolerate cannabinoids poorly.

Dyspnea

FREQUENCY

Dyspnea is the subjective experience of being short of breath. Over 50%, and as many as 75%, of dying patients, especially those with lung cancer, metastases to the lung, CHF, and COPD, experience dyspnea at some point near the end of life. Dyspnea is among the most distressing of physical symptoms and can be even more distressing than pain.

ASSESSMENT

As with pain, dyspnea is a subjective experience that may not correlate with objective measures of Po_2, Pco_2, or respiratory rate. Consequently, measurements of oxygen saturation through pulse oximetry or blood gases are rarely helpful in guiding therapy. Despite the limitations of existing assessment methods, physicians should regularly assess and document patients’ experience of dyspnea and its intensity. Guidelines recommend visual analogue dyspnea scales to assess the severity of symptoms and the effects of treatment. Potentially reversible or treatable causes of dyspnea include infection, pleural effusions, pulmonary emboli, pulmonary edema, asthma, and tumor encroachment on the airway. However, the risk-versus-benefit ratio of the diagnostic and therapeutic interventions for patients with little time left to live must be considered carefully before undertaking diagnostic steps. Frequently, the specific etiology cannot be identified, and dyspnea is the consequence of progression of the underlying disease that cannot be treated. The anxiety caused by dyspnea and the choking sensation can significantly exacerbate the underlying dyspnea in a negatively reinforcing cycle.

INTERVENTIONS

When reversible or treatable etiologies are diagnosed, they should be treated as long as the side effects of treatment, such as repeated drainage of

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Usually, treatment will be symptomatic **(Table 12-6)**. Supplemental oxygen does not appear to be effective. “A systematic review of the literature failed to demonstrate a consistent beneficial effect of oxygen inhalation over air inhalation for study participants with dyspnea due to end-stage cancer

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

When reversible or treatable etiologies are diagnosed, they should be treated as long as the side effects of treatment, such as repeated drainage of effusions or anticoagulants, are less burdensome than the dyspnea itself. More aggressive treatments such as stenting a bronchial lesion may be warranted if it is clear that the dyspnea is due to tumor invasion at that site and if the patient and family understand the risks of such a procedure.

Usually, treatment will be symptomatic **(Table 12-6)**. Supplemental oxygen does not appear to be effective. "A systematic review of the literature failed to demonstrate a consistent beneficial effect of oxygen inhalation over air inhalation for study participants with dyspnea due to end-stage cancer or cardiac failure." Therefore, oxygen may be no more than an expensive placebo. Low-dose opioids reduce the sensitivity of the central respiratory center and relieve the sensation of dyspnea. If patients are not receiving opioids, weak opioids can be initiated; if patients are already receiving opioids, morphine or other stronger opioids should be used. Controlled trials do not support the use of nebulized opioids for dyspnea at the end of life. Phenothiazines and chlorpromazine may be helpful when combined with opioids. Benzodiazepines can be helpful in treating dyspnea, but only if anxiety is present. Benzodiazepines should not be used as first-line therapy or if there is no anxiety. If the patient has a history of COPD or asthma, inhaled bronchodilators and glucocorticoids may be helpful. If the patient has pulmonary edema due to heart failure, diuresis with a medication such as furosemide is indicated. Excess secretions can be transdermally or intravenously dried with scopolamine. More general interventions that medical staff can perform include sitting the patient upright, removing smoke or other irritants like perfume, ensuring a supply of fresh air with sufficient humidity, and minimizing other factors that can increase anxiety.

TABLE 12-6

Medications for the Management of Dyspnea

INTERVENTION	DOSE	COMMENTS
Weak opioids		For patients with mild dyspnea
Codeine (or codeine with 325 mg acetaminophen)	30 mg PO q4h	For opioid-naïve patients
Hydrocodone	5 mg PO q4h	
Strong opioids		For opioid-naïve patients with moderate to severe dyspnea
Morphine	5–10 mg PO q4h	For patients already taking opioids for pain or other symptoms
	30–50% of baseline opioid dose q4h	
Oxycodone	5–10 mg PO q4h	
Hydromorphone	1–2 mg PO q4h	
Anxiolytics		Give a dose every hour until the patient is relaxed; then provide a dose for maintenance
Lorazepam	0.5–2.0 mg PO/SL/IV qh then q4–6h	
Clonazepam	0.25–2.0 mg PO q12h	
Midazolam	0.5 mg IV q15min	

Fatigue

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Fatigue is one of the most commonly reported symptoms not only of cancer treatment but also of the palliative care of multiple sclerosis, COPD, heart failure, and HIV. More than 90% of terminally ill patients experience fatigue and/or weakness. Fatigue is frequently cited as one of the most distressing

Fatigue

FREQUENCY

Fatigue is one of the most commonly reported symptoms not only of cancer treatment but also of the palliative care of multiple sclerosis, COPD, heart failure, and HIV. More than 90% of terminally ill patients experience fatigue and/or weakness. Fatigue is frequently cited as one of the most distressing symptoms in these patients.

ETIOLOGY

The multiple causes of fatigue in the terminally ill can be categorized as resulting from the underlying disease; from disease-induced factors such as tumor necrosis factor and other cytokines; and from secondary factors such as dehydration, anemia, infection, hypothyroidism, and drug side effects. In addition to low caloric intake, loss of muscle mass and changes in muscle enzymes may play an important role in fatigue during terminal illness. The importance of changes in the CNS, especially the reticular activating system, have been hypothesized based on reports of fatigue in patients receiving cranial radiation, experiencing depression, or having chronic pain in the absence of cachexia or other physiologic changes. Finally, depression and other causes of psychological distress can contribute to fatigue.

ASSESSMENT

Like pain and dyspnea, fatigue is subjective, as it represents a patient's sense of tiredness and decreased capacity for physical work. Objective changes, even in body mass, may be absent. Consequently, assessment must rely on patient self-reporting. Scales used to measure fatigue, such as the Edmonton Functional Assessment Tool, the Fatigue Self-Report Scales, and the Rhoten Fatigue Scale, are usually appropriate for research but not clinical purposes. In clinical practice, a simple performance assessment such as the Karnofsky performance status or the Eastern Cooperative Oncology Group (ECOG)'s question "How much of the day does the patient spend in bed?" may be the best measure. In the ECOG 0–4 performance status assessment, 0 = normal activity; 1 = symptomatic without being bedridden; 2 = requiring some, but <50%, bed time; 3 = bedbound more than half the day; and 4 = bedbound all the time. Such a scale allows for assessment over time and correlates with overall disease severity and prognosis. A 2008 review by the European Association of Palliative Care also described several longer assessment tools that contained 9–20 items, including the Piper Fatigue Inventory, the Multidimensional Fatigue Inventory, and the Brief Fatigue Inventory (BFI).

INTERVENTIONS

Reversible causes of fatigue, such as anemia and infection, should be treated. However, at the end of life, it must be realistically acknowledged that fatigue will not be "cured." The goal is to ameliorate fatigue and help patients and families adjust expectations. Behavioral interventions should be utilized to avoid blaming the patient for inactivity and to educate both the family and the patient that the underlying disease causes physiologic changes that produce low energy levels. Understanding that the problem is physiologic and not psychological can help alter expectations regarding the patient's level of physical activity. Practically, this may mean reducing routine activities such as housework, cooking, and social events outside the house and making it acceptable to receive guests while lying on a couch. At the same time, the implementation of exercise regimens and physical therapy can raise endorphins, reduce muscle wasting, and decrease the risk of depression. In addition, ensuring good hydration without worsening edema may help reduce fatigue. Discontinuing medications that worsen fatigue may help, including cardiac medications, benzodiazepines, certain antidepressants, or opioids if the patient's pain is well-controlled. As end-of-life care proceeds into its final stages, fatigue may protect patients from further suffering, and continued treatment could be detrimental.

Only a few pharmacologic interventions target fatigue and weakness. Randomized controlled trials suggest glucocorticoids can increase energy and enhance mood. Dexamethasone (8 mg/d) is preferred for its once-a-day dosing and minimal mineralocorticoid activity. Benefit, if any, is usually seen within the first month. For fatigue related to anorexia, megestrol (480–800 mg) can be helpful. Psychostimulants such as dextroamphetamine (5–10 mg PO) and methylphenidate (2.5–5 mg PO) may enhance energy levels, although controlled trials have not shown these drugs to be effective for fatigue induced by mild to moderate cancer. Doses should be given in the morning and at noon to minimize the risk of counterproductive insomnia. Modafinil and armodafinil, developed for narcolepsy, have shown promise in the treatment of fatigue and have the advantage of once-daily dosing. Their precise role in fatigue at the end of life has not been documented but may be worth trying if other interventions are not beneficial. Anecdotal evidence suggests that L-carnitine may improve fatigue, depression, and sleep disruption.

PALLIATIVE SEDATION

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

done at the request of the family rather than according to the patient's wishes, or when there are other interventions that could still be tried. The use of

PALLIATIVE SEDATION

Palliative sedation is used in distressing situations that cannot be addressed in other ways. When patients experience severe symptoms, such as pain or dyspnea, that cannot be relieved by conventional interventions or experience acute catastrophic symptoms, such as uncontrolled seizures, then palliative sedation should be considered as an intervention of last resort. It can be abused if done to hasten death (which it usually does not), when done at the request of the family rather than according to the patient's wishes, or when there are other interventions that could still be tried. The use of palliative sedation in cases of extreme existential or spiritual distress remains controversial. Typically, palliative sedation should be introduced only after the patient and family have been assured that all other interventions have been tried and after the patient and their loved ones have been able to "say goodbye."

Palliative sedation can be achieved by significantly increasing opioid doses until patients become unconscious and then putting them on a continuous infusion. Another commonly used medication for palliative sedation is midazolam at 1–5 mg IV every 5–15 min to calm the patient, followed by a continuous IV or subcutaneous infusion of 1 mg/h. In hospital settings, a continuous propofol infusion of 5 μg/kg per min can be used. There are also other, less commonly used medications for palliative sedation that include levomepromazine, chlorpromazine, and phenobarbital.

PSYCHOLOGICAL SYMPTOMS AND THEIR MANAGEMENT

Depression

FREQUENCY AND IMPACT

Depression at the end of life presents an apparently paradoxical situation. Many people believe that depression is normal among seriously ill patients because they are dying. People frequently say, "Wouldn't you be depressed?" Although sadness, anxiety, anger, and irritability are normal responses to a serious condition, they are typically of modest intensity and transient. Persistent sadness and anxiety and the physically disabling symptoms that they can lead to are abnormal and suggestive of major depression. The precise number of terminally ill patients who are depressed is uncertain, primarily due to a lack of consistent diagnostic criteria and screening. Careful follow-up of patients suggests that while as many as 75% of terminally ill patients experience depressive symptoms, ~25% of terminally ill patients have major depression. Depression at the end of life is concerning because it can decrease the quality of life, interfere with closure in relationships and other separation work, obstruct adherence to medical interventions, and amplify the suffering associated with pain and other symptoms.

ETIOLOGY

Previous history of depression, family history of depression or bipolar disorder, and prior suicide attempts are associated with increased risk for depression among terminally ill patients. Other symptoms, such as pain and fatigue, are associated with higher rates of depression; uncontrolled pain can exacerbate depression, and depression can cause patients to be more distressed by pain. Many medications used in the terminal stages, including glucocorticoids, and some anticancer agents, such as tamoxifen, interleukin 2, interferon α, and vincristine, also are associated with depression. Some terminal conditions, such as pancreatic cancer, certain strokes, and heart failure, have been reported to be associated with higher rates of depression, although this is controversial. Finally, depression may be attributable to grief over the loss of a role or function, social isolation, or loneliness.

ASSESSMENT

Unfortunately, many studies suggest that most depressed patients at the end of life are not diagnosed, or if they are diagnosed, they are not properly treated. Diagnosing depression among seriously ill patients is complicated, as many of the vegetative symptoms in the *DSM-V* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) criteria for clinical depression—insomnia, anorexia and weight loss, fatigue, decreased libido, and difficulty concentrating—are associated with the process of dying itself. The assessment of depression in seriously ill patients therefore should focus on the dysphoric mood, helplessness, hopelessness, and lack of interest, enjoyment, and concentration in normal activities. It is now recommended that patients near the end of life should be screened with either the PHQ-9 or the PHQ-2, which asks "Over the past 2 weeks, how often have you been bothered by any of the following problems? (1) Little interest or pleasure in doing things and (2) feeling down, depressed or hopeless." The answer categories are as follows: not at all, several days, more than half the days, nearly every day. Other possible diagnostic tools include the short form of the Beck Depression Index or a visual analogue scale.

Certain conditions may be confused with depression. Endocrinopathies, such as hypothyroidism and Cushing's syndrome, electrolyte abnormalities, such as hypercalcemia, and akathisia, especially from dopamine-blocking antiemetics such as metoclopramide and prochlorperazine, can mimic depression and should be excluded.

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Undertreatment of depressed, terminally ill patients is common. Physicians must treat any physical symptom, such as pain, that may be causing or exacerbating depression. Fostering adaptation to the many losses that the patient is experiencing can also be helpful. Unfortunately, there are few

such as hypercalcemia, and akathisia, especially from dopamine-blocking antiemetics such as metoclopramide and prochlorperazine, can mimic depression and should be excluded.

INTERVENTIONS

Undertreatment of depressed, terminally ill patients is common. Physicians must treat any physical symptom, such as pain, that may be causing or exacerbating depression. Fostering adaptation to the many losses that the patient is experiencing can also be helpful. Unfortunately, there are few randomized trials to guide such interventions. Thus, treatment typically follows the treatment used for non–terminally ill depressed patients.

In the absence of randomized controlled trials, nonpharmacologic interventions, including group or individual psychological counseling, and behavioral therapies such as relaxation and imagery can be helpful, especially in combination with drug therapy.

Pharmacologic interventions remain at the core of therapy. The same medications are used to treat depression in terminally ill as in non–terminally ill patients. Psychostimulants may be preferred for patients with a poor prognosis or for those with fatigue or opioid-induced somnolence. Psychostimulants are comparatively fast-acting, working within a few days instead of the weeks required for selective serotonin reuptake inhibitors (SSRIs). Dextroamphetamine or methylphenidate should be started at 2.5–5.0 mg in the morning and at noon, the same starting doses used for treating fatigue. The doses can eventually be escalated up to 15 mg bid. Modafinil is started at 100 mg qd and can be increased to 200 mg if there is no effect at the lower dose. Pemoline is a nonamphetamine psychostimulant with minimal abuse potential. It is also effective as an antidepressant beginning at 18.75 mg in the morning and at noon. Because it can be absorbed through the buccal mucosa, it is preferred for patients with intestinal obstruction or dysphagia. If it is used for prolonged periods, liver function must be monitored. The psychostimulants can also be combined with more traditional antidepressants while waiting for the antidepressants to become effective, then tapered down after a few weeks if necessary. Psychostimulants have side effects, particularly initial anxiety, insomnia, and very rarely paranoia, which may necessitate lowering the dose or discontinuing treatment.

Mirtazapine, an antagonist at the postsynaptic serotonin receptors, is a promising psychostimulant. It should be started at 7.5 mg before bed and titrated up no more than once every 1–2 weeks to a maximal dose of 45 mg/d. It has sedating, antiemetic, and anxiolytic properties, with few drug interactions. Its side effect of weight gain may be beneficial for seriously ill patients; it is available in orally disintegrating tablets.

For patients with a prognosis of several months or longer, SSRIs, including fluoxetine, sertraline, paroxetine, escitalopram, and citalopram, and serotonin-noradrenaline reuptake inhibitors, such as venlafaxine and duloxetine, are the preferred treatments, due to their efficacy and comparatively few side effects. Because low doses of these medications may be effective for seriously ill patients, one should use half the usual starting dose as for healthy adults. The starting dose for fluoxetine is 10 mg once a day. In most cases, once-a-day dosing is possible. The choice of which SSRI to use should be driven by (1) the patient's past success or failure with the specific medication and (2) the most favorable side effect profile for that specific agent. For instance, for a patient in whom fatigue is a major symptom, a more activating SSRI (fluoxetine) would be appropriate. For a patient in whom anxiety and sleeplessness are major symptoms, a more sedating SSRI (paroxetine) would be appropriate. Importantly, it can take up to 4 weeks for these drugs to have an effect.

Atypical antidepressants are recommended only in select circumstances, usually with the assistance of a specialty consultation. Trazodone can be an effective antidepressant but is sedating and can cause orthostatic hypotension and, occasionally, priapism. Therefore, it should be used before bed and only when a sedating effect is desired and is often used for patients with insomnia at a dose starting at 25 mg. Bupropion can also be used. In addition to its antidepressant effects, bupropion is energizing, making it useful for depressed patients who experience fatigue. However, it can cause seizures, preventing its use for patients with a risk of CNS neoplasms or terminal delirium. Finally, alprazolam, a benzodiazepine, starting at 0.25–1.0 mg tid, can be effective in seriously ill patients who have a combination of anxiety and depression. Although it is potent and works quickly, it has many drug interactions and may cause delirium, especially among very ill patients, because of its strong binding to the benzodiazepine–γ-aminobutyric acid (GABA) receptor complex.

Unless used as adjuvants for the treatment of pain, tricyclic antidepressants are not recommended. While they can be effective, their therapeutic window and serious side effects typically limit their utility. Similarly, monoamine oxidase (MAO) inhibitors are not recommended because of their side effects and dangerous drug interactions.

Delirium

(See Chap. 27)

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

common in the days and hours immediately before death. Up to 85% of patients dying from cancer may experience terminal delirium.

FREQUENCY

In the weeks or months before death, delirium is uncommon, although it may be significantly underdiagnosed. However, delirium becomes relatively common in the days and hours immediately before death. Up to 85% of patients dying from cancer may experience terminal delirium.

ETIOLOGY

Delirium is a global cerebral dysfunction characterized by alterations in cognition and consciousness. It is frequently preceded by anxiety, changes in sleep patterns (especially reversal of day and night), and decreased attention. In contrast to dementia, delirium has an acute onset, is characterized by fluctuating consciousness and inattention, and is reversible, although reversibility may be more theoretical than real for patients near death. Delirium may occur in a patient with dementia; indeed, patients with dementia are more vulnerable to delirium.

Causes of delirium include metabolic encephalopathy arising from liver or renal failure, hypoxemia, or infection; electrolyte imbalances such as hypercalcemia; paraneoplastic syndromes; dehydration; and primary brain tumors, brain metastases, or leptomeningeal spread of tumor. Among dying patients, delirium is commonly caused by side effects of treatments, including radiation for brain metastases and medications, such as opioids, glucocorticoids, anticholinergic drugs, antihistamines, antiemetics, benzodiazepines, and chemotherapeutic agents. The etiology may be multifactorial; e.g., dehydration may exacerbate opioid-induced delirium.

ASSESSMENT

Delirium should be recognized in any terminally ill patient exhibiting new onset of disorientation, impaired cognition, somnolence, fluctuating levels of consciousness, or delusions with or without agitation. Delirium must be distinguished from acute anxiety, depression, and dementia. The central distinguishing feature is altered consciousness, which usually is not noted in anxiety, depression, or dementia. Although “hyperactive” delirium, characterized by overt confusion and agitation, is probably more common, patients should also be assessed for “hypoactive” delirium, which is characterized by sleep-wake reversal and decreased alertness.

In some cases, use of formal assessment tools such as the Mini-Mental Status Examination (which does not distinguish delirium from dementia) and the Delirium Rating Scale (which does distinguish delirium from dementia) may be helpful in distinguishing delirium from other processes. The patient’s list of medications must be evaluated carefully. Nonetheless, a reversible etiologic factor for delirium is found in fewer than half of all terminally ill patients. Given that most terminally ill patients experiencing delirium are very close to death and often at home, extensive diagnostic evaluations such as lumbar punctures and neuroradiologic examinations are inappropriate.

INTERVENTIONS

One of the most important objectives of terminal care is to provide terminally ill patients the lucidity to say goodbye to the people they love. Delirium, especially when in combination with agitation during the final days, is distressing to family and caregivers. A strong determinant of bereavement difficulties is witnessing a difficult death. Thus, terminal delirium should be treated aggressively.

At the first sign of delirium, such as day-night reversal with slight changes in mentation, the physician should let the family members know that it is time to be sure that everything they want to say has been said. The family should be informed that delirium is common just before death.

If medications are suspected of being a cause of the delirium, unnecessary agents should be discontinued. Other potentially reversible causes, such as constipation, urinary retention, and metabolic abnormalities, should be treated. Supportive measures aimed at providing a familiar environment should be instituted, including restricting visits only to individuals with whom the patient is familiar and eliminating new experiences; orienting the patient, if possible, by providing a clock and calendar; and gently correcting the patient’s hallucinations or cognitive mistakes.

Pharmacologic management focuses on the use of neuroleptics and, in extreme cases, anesthetics **(Table 12-7)**. Haloperidol remains the first-line therapy. Usually, patients can be controlled with a low dose (1–3 mg/d), given every 6 h, although some may require as much as 20 mg/d. Haloperidol can be administered PO, SC, or IV. IM injections should not be used, except when this is the only way to address a patient’s delirium. Olanzapine, an atypical neuroleptic, has shown significant effectiveness in completely resolving delirium in cancer patients. It also has other beneficial effects for terminally ill patients, including antinausea, antianxiety, and weight gain. Olanzapine is useful for patients with longer anticipated life expectancies because it is less likely to cause dysphoria and has a lower risk of dystonic reactions. Additionally, because olanzapine is metabolized through multiple pathways, it can be used in patients with hepatic and renal dysfunction. Olanzapine has the disadvantage that it is only available orally and takes a

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Neuroleptics may be combined with lorazepam to reduce agitation when the delirium is the result of alcohol or sedative withdrawal.

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

therapy. Usually, patients can be controlled with a low dose (1–3 mg/d), given every 6 h, although some may require as much as 20 mg/d. Haloperidol because it is less likely to cause dysphoria and has a lower risk of dystonic reactions. Additionally, because olanzapine is metabolized through multiple pathways, it can be used in patients with hepatic and renal dysfunction. Olanzapine has the disadvantage that it is only available orally and takes a week to reach steady state. The usual dose is 2.5–5 mg PO bid. Chlorpromazine (10–25 mg every 4–6 h) can be useful if sedation is desired and can be administered IV or PR in addition to PO. Dystonic reactions resulting from dopamine blockade are a side effect of neuroleptics, although they are reported to be rare when these drugs are used to treat terminal delirium. If patients develop dystonic reactions, benztropine should be administered. Neuroleptics may be combined with lorazepam to reduce agitation when the delirium is the result of alcohol or sedative withdrawal.

TABLE 12-7

Medications for the Management of Delirium

INTERVENTIONS	DOSE
Neuroleptics	
Haloperidol	0.5–5 mg q2–12h, PO/IV/SC/IM
Thioridazine	10–75 mg q4–8h, PO
Chlorpromazine	12.5–50 mg q4–12h, PO/IV/IM
Atypical neuroleptics	
Olanzapine	2.5–5 mg qd or bid, PO
Risperidone	1–3 mg q12h, PO
Anxiolytics	
Lorazepam	0.5–2 mg q1–4h, PO/IV/IM
Midazolam	1–5 mg/h continuous infusion, IV/SC
Anesthetics	
Propofol	0.3–2.0 mg/h continuous infusion, IV

If no response to first-line therapy is observed, a specialty consultation should be obtained with a goal to change to a different medication. If the patient fails to improve after a second neuroleptic, sedation with either an anesthetic such as propofol or continuous-infusion midazolam may be necessary. By some estimates, as many as 25% of patients at the very end of life who experience delirium, especially restless delirium with myoclonus or convulsions, may require sedation.

Physical restraints should be used with great reluctance and only when patients' violence is threatening to themselves or others. If restraints are used, their appropriateness should be frequently reevaluated.

Insomnia

FREQUENCY

Sleep disorders, defined as difficulty initiating sleep or maintaining sleep, sleep difficulty at least 3 nights a week, or sleep difficulty that causes impairment of daytime functioning, occurs in 19–63% of patients with advanced cancer. Some 30–74% of patients with other end-stage conditions, including AIDS, heart disease, COPD, and renal disease, experience insomnia.

ETIOLOGY

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

glucocorticoids, and β agonists, are significant contributors to sleep disorders, as are caffeine and alcohol. Multiple over-the-counter medications contain caffeine and antihistamines, which can contribute to sleep disorders.

ETIOLOGY

Patients with cancer may experience changes in sleep efficiency, such as an increase in stage I sleep. Insomnia may also coexist with both physical illnesses, like thyroid disease, and psychological illnesses, like depression and anxiety. Medications, including antidepressants, psychostimulants, glucocorticoids, and β agonists, are significant contributors to sleep disorders, as are caffeine and alcohol. Multiple over-the-counter medications contain caffeine and antihistamines, which can contribute to sleep disorders.

ASSESSMENT

Assessments should include specific questions concerning sleep onset, sleep maintenance, and early-morning wakening, as these will provide clues to both the causative agents and management of insomnia. Patients should be asked about previous sleep problems, screened for depression and anxiety, and asked about symptoms of thyroid disease. Caffeine and alcohol are prominent causes of sleep problems, and a careful history of the use of these substances should be obtained. Both excessive use and withdrawal from alcohol can be causes of sleep problems.

INTERVENTIONS

The mainstays of any intervention include improvement of sleep hygiene (encouragement of regular time for sleep, decreased nighttime distractions, elimination of caffeine and other stimulants and alcohol), interventions to treat anxiety and depression, and treatment for the insomnia itself. For patients with depression who have insomnia and anxiety, a sedating antidepressant such as mirtazapine can be helpful. In the elderly, trazodone, beginning at 25 mg at nighttime, is an effective sleep aid at doses lower than those that cause its antidepressant effect. Zolpidem may have a decreased incidence of delirium in patients compared with traditional benzodiazepines, but this has not been clearly established. When benzodiazepines are prescribed, short-acting ones (such as lorazepam) are favored over longer-acting ones (such as diazepam). Patients who receive these medications should be observed for signs of increased confusion and delirium.

SOCIAL NEEDS AND THEIR MANAGEMENT

Financial Burdens

FREQUENCY

Dying can impose substantial economic strains on patients and families, potentially causing distress. This is known as financial toxicity. In the United States, which has the least comprehensive health insurance systems among wealthy countries, a quarter of families coping with end-stage cancer report that care was a major financial burden and a third used up most of their savings. Among Medicare beneficiaries, average out-of-pocket costs were >$8000. Between 10% and 30% of families are forced to sell assets, use savings, or take out a mortgage to pay for the patient's health care costs.

The patient is likely to reduce hours worked and eventually stop working altogether. In 20% of cases, a family member of the terminally ill patient also must stop working to provide care. The major underlying causes of economic burden are related to poor physical functioning and care needs, such as the need for housekeeping, nursing, and personal care. More debilitated patients and poor patients experience greater economic burdens.

INTERVENTION

The economic burden of end-of-life care should not be ignored as a private matter. It has been associated with a number of adverse health outcomes, including preferring comfort care over life-prolonging care, as well as consideration of euthanasia or physician-assisted suicide (PAS). Economic burdens increase the psychological distress of the families and caregivers of terminally ill patients, and poverty is associated with many adverse health outcomes. Importantly, studies have found that "patients with advanced cancer who reported having end-of-life conversations with physicians had significantly lower health care costs in their final week of life. Higher costs were associated with worse quality of death." Assistance from a social worker, early on if possible, to ensure access to all available benefits may be helpful. Many patients, families, and health care providers are unaware of options for long-term care insurance, respite care, the Family Medical Leave Act (FMLA), and other sources of assistance. Some of these options (such as respite care) may be part of a formal hospice program, but others (such as the FMLA) do not require enrollment in a hospice program.

Relationships

FREQUENCY

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

INTERVENTIONS

Settling personal issues and closing the narrative of lived relationships are universal needs. When asked if sudden death or death after an illness is preferable, respondents often initially select the former, but soon change to the latter as they reflect on the importance of saying goodbye. Bereaved family members who have not had the chance to say goodbye often have a more difficult grief process.

INTERVENTIONS

Care of seriously ill patients requires efforts to facilitate the types of encounters and time spent with family and friends that are necessary to meet those needs. Family and close friends may need to be accommodated in hospitals and other facilities with unrestricted visiting hours, which may include sleeping near the patient, even in otherwise regimented institutional settings. Physicians and other health care providers may be able to facilitate and resolve strained interactions between the patient and other family members. Assistance for patients and family members who are unsure about how to create or help preserve memories, whether by providing materials such as a scrapbook or memory box or by offering them suggestions and informational resources, can be deeply appreciated. Taking photographs and creating videos can be especially helpful to terminally ill patients who have younger children or grandchildren.

Family Caregivers

FREQUENCY

Caring for seriously ill patients places a heavy burden on families. Families are frequently required to provide transportation and homemaking, as well as other services. Typically, paid professionals, such as home health nurses and hospice workers, supplement family care; only about a quarter of all caregiving consists of exclusively paid professional assistance. Over the past 40 years, there has been a significant decline in the United States of deaths occurring in hospitals, with a simultaneous increase in deaths in other facilities and at home. Over a third of deaths occur in patients' homes. This increase in out-of-hospital deaths increases reliance on families for end-of-life care. Increasingly, family members are being called upon to provide physical care (such as moving and bathing patients) and medical care (such as assessing symptoms and giving medications) in addition to emotional care and support.

Three-quarters of family caregivers of terminally ill patients are women—wives, daughters, sisters, and even daughters-in-law. Since many are widowed, women tend to be able to rely less on family for caregiving assistance and may need more paid assistance. About 20% of terminally ill patients report substantial unmet needs for nursing and personal care. The impact of caregiving on family caregivers is substantial: both bereaved and current caregivers have a higher mortality rate than that of non-caregiving controls.

INTERVENTIONS

It is imperative to inquire about unmet needs and to try to ensure that those needs are met either through the family or by paid professional services when possible. Community assistance through houses of worship or other community groups often can be mobilized by telephone calls from the medical team to someone the patient or family identifies. Sources of support specifically for family caregivers should be identified through local sources or nationally through groups such as the National Family Caregivers Association (*www.nfcacares.org*), the American Cancer Society (*www.cancer.org*), and the Alzheimer's Association (*www.alz.org*).

EXISTENTIAL NEEDS AND THEIR MANAGEMENT

Frequency

Religion and spirituality are often important to dying patients. Nearly 70% of patients report becoming more religious or spiritual when they became terminally ill, and many find comfort in religious or spiritual practices such as prayer. However, ~20% of terminally ill patients become less religious, frequently feeling cheated or betrayed by becoming terminally ill. For other patients, the need is for existential meaning and purpose that is distinct from, and may even be antithetical to, religion or spirituality. When asked, patients and family caregivers frequently report wanting their professional caregivers to be more attentive to religion and spirituality.

Assessment

Health care providers are often hesitant about involving themselves in the religious, spiritual, and existential experiences of their patients because it may seem private or not relevant to the current illness. But physicians and other members of the care team should be able at least to detect spiritual

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

suitable. Pastoral care providers may be helpful, whether from the medical institution or from the patient's own community.

Health care providers are often hesitant about involving themselves in the religious, spiritual, and existential experiences of their patients because it may seem private or not relevant to the current illness. But physicians and other members of the care team should be able at least to detect spiritual and existential needs. Screening questions have been developed for a physician's spiritual history taking. Spiritual distress can amplify other types of suffering and even masquerade as intractable physical pain, anxiety, or depression. The screening questions in the comprehensive assessment are usually sufficient. Deeper evaluation and intervention are rarely appropriate for the physician unless no other member of a care team is available or suitable. Pastoral care providers may be helpful, whether from the medical institution or from the patient's own community.

Interventions

Precisely how religious practices, spirituality, and existential explorations can be facilitated and improve end-of-life care is not well established. What is clear is that for physicians, one main intervention is to inquire about the role and importance of spirituality and religion in a patient's life. This will help a patient feel heard and help physicians identify specific needs. In one study, only 36% of respondents indicated that a clergy member would be comforting. Nevertheless, the increase in religious and spiritual interest among a substantial fraction of dying patients suggests inquiring of individual patients how this need can be addressed. Some evidence supports specific methods of addressing existential needs in patients, ranging from establishing a supportive group environment for terminal patients to individual treatments emphasizing a patient's dignity and sources of meaning.

MANAGING THE LAST STAGES

PALLIATIVE CARE SERVICES: HOW AND WHERE

Determining the best approach to providing palliative care to patients will depend on patient preferences, the availability of caregivers and specialized services in close proximity, institutional resources, and reimbursement. Hospice is a leading, but not the only, model of palliative care services. In the United States, slightly more than a third—35.7%—of hospice care is provided in private residential homes with 14.5% of hospice care in nursing homes. In the United States, Medicare pays for hospice services under Part A, the hospital insurance part of reimbursement. Two physicians must certify that the patient has a prognosis of ≤6 months if the disease runs its usual course. Prognoses are probabilistic by their nature; patients are not required to die within 6 months but rather to have a condition from which half the individuals with it would not be alive within 6 months. Patients sign a hospice enrollment form that states their intent to forgo curative services related to their terminal illness but can still receive medical services for other comorbid conditions. Patients also can withdraw enrollment and reenroll later; the hospice Medicare benefit can be revoked later to secure traditional Medicare benefits. Payments to the hospice are per diem (or capitated), not fee-for-service. Payments are intended to cover physician services for the medical direction of the care team; regular home care visits by registered nurses and licensed practical nurses; home health aide and homemaker services; chaplain services; social work services; bereavement counseling; and medical equipment, supplies, and medications. No specific therapy is excluded, and the goal is for each therapy to be considered for its symptomatic (as opposed to disease-modifying) effect. Additional clinical care, including services of the primary physician, is covered by Medicare Part B even while the hospice Medicare benefit is in place.

The Affordable Care Act directs the secretary of Health and Human Services to gather data on Medicare hospice reimbursement with the goal of reforming payment rates to account for resource use over an entire episode of care. The legislation also requires additional evaluations and reviews of eligibility for hospice care by hospice physicians or nurses. The Center for Medicare and Medicaid Innovation (CMMI) sponsors and carries out demonstration projects to test models and evaluate the potential of new methods. In 2016, CMMI started a 5-year test of concurrent hospice and palliative care services with curative treatment for terminally ill patients who have a life expectancy of ≤6 months. A 4-year test initiated in 2021 will examine the inclusion of hospice in Medicare Advantage covering 8% of the market and include important health plans.

By 2018, the average length of enrollment in a hospice for Medicare beneficiaries was 90 days. However, the median length of stay was just 18 days, suggesting most patients are in hospice for a short time. Such short stays create barriers to establishing high-quality palliative services in patients' homes and also place financial strains on hospice providers since the initial assessments are resource intensive. Physicians should initiate early referrals to the hospice to allow more time for patients to receive palliative care.

In the United States, hospice care has been the main method for securing palliative services for terminally ill patients. However, leading physicians have increasingly emphasized the need to introduce palliative care much earlier in patients' illness, and efforts are being made to develop palliative care services that can be provided before the last 6 months of life and across a variety of settings. Studies of terminally ill patients indicate that those who received in-home palliative care delivered by an interdisciplinary team compared to usual care were more satisfied, more likely to die at home, and had fewer visits to the emergency room and lower per-day costs. More companies and home health agencies are now offering nonhospice palliative care services in patients' homes in an effort to increase quality of life and forestall emergency room visits and hospitalizations. Similarly, facilitating a smoother transition to hospice care closer to death.

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

palliative care services in patients' homes in an effort to increase quality of life and forestall emergency room visits and hospitalizations. Similarly, palliative care services are increasingly available via consultation, rather than being available only in hospital, day care, outpatient, and nursing home settings. Palliative care consultations for nonhospice patients can be billed as for other consultations under Medicare Part B. It is argued that using palliative care earlier in patients' illness allows patients and family members to become more acculturated to avoiding life-sustaining treatments, facilitating a smoother transition to hospice care closer to death.

WITHDRAWING AND WITHHOLDING LIFE-SUSTAINING TREATMENT

Legal Aspects

For centuries, it has been deemed ethical to withhold or withdraw life-sustaining interventions. The current legal consensus in the United States and most wealthy countries is that patients have a moral as well as legal right to refuse medical interventions. American courts also have held that incompetent patients have a right to refuse medical interventions. For patients who are incompetent and terminally ill and who have not completed an advance care directive, next of kin can exercise that right, although this may be restricted in some states, depending on how clear and convincing the evidence is of the patient's preferences. Courts have limited families' ability to terminate life-sustaining treatments in patients who are conscious and incompetent but not terminally ill. In theory, patients' right to refuse medical therapy can be limited by four countervailing interests: (1) preservation of life, (2) prevention of suicide, (3) protection of third parties such as children, and (4) preservation of the integrity of the medical profession. In practice, these interests almost never override the right of competent patients and incompetent patients who have left explicit wishes or advance care directives.

For incompetent patients who either appointed a proxy without specific indications of their wishes or never completed an advance care directive, three criteria have been suggested to guide the decision to terminate medical interventions. First, some commentators suggest that ordinary care should be administered but extraordinary care could be terminated. Because the ordinary/extraordinary distinction is too vague, courts and commentators widely agree that it should not be used to justify decisions about stopping treatment. Second, many courts have advocated the use of the substituted-judgment criterion, which holds that the proxy decision-makers should try to imagine what the incompetent patient would do if he or she were competent. However, multiple studies indicate that many proxies, even close family members, cannot accurately predict what the patient would have wanted. Therefore, substituted judgment becomes more of a guessing game than a way of fulfilling the patient's wishes. Finally, the best-interests criterion holds that proxies should evaluate treatments by balancing their benefits and risks and select those treatments where the benefits maximally outweigh the burdens of treatment. Clinicians have a clear and crucial role in this by carefully and dispassionately explaining the known benefits and burdens of specific treatments. Yet even when that information is as clear as possible, different individuals can have very different views of what is in the patient's best interests, and families may have disagreements or even overt conflicts. This criterion has been criticized because there is no single way to determine the balance between benefits and burdens; it depends on a patient's personal values. For instance, for some people, being alive even if mentally incapacitated is a benefit, whereas for others, it may be the worst possible existence. As a matter of practice, physicians rely on family members to make decisions that they feel are best and object only if those decisions seem to demand treatments that the physicians consider not beneficial.

Practices

Withholding and withdrawing acutely life-sustaining medical interventions from terminally ill patients are now standard practice. More than 90% of American patients die without cardiopulmonary resuscitation (CPR), and just as many forgo other potentially life-sustaining interventions. For instance, in ICUs in the period of 1987–1988, CPR was performed 49% of the time, but it was performed only 10% of the time in 1992–1993 and on just 1.8% of admissions from 2001 to 2008. On average, 3.8 interventions, such as vasopressors and transfusions, were stopped for each dying ICU patient. However, up to 19% of decedents in hospitals received interventions such as extubation, ventilation, and surgery in the 48 h preceding death. There is wide variation in practices among hospitals and ICUs, suggesting an important element of physician preferences rather than consistent adherence to professional society recommendations.

Mechanical ventilation may be the most challenging intervention to withdraw. The two approaches are *terminal extubation*, which is the removal of the endotracheal tube, and *terminal weaning*, which is the gradual reduction of the fraction of inspired oxygen (FIO_2) or ventilator rate. One-third of ICU physicians prefer to use the terminal weaning technique, and 13% extubate; the majority of physicians utilize both techniques. The American Thoracic Society's 2008 clinical policy guidelines note that there is no single correct process of ventilator withdrawal and that physicians use and should be proficient in both methods but that the chosen approach should carefully balance benefits and burdens as well as patient and caregiver preferences. Some recommend terminal weaning because patients do not develop upper airway obstruction and the distress caused by secretions or stridor;

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

the titration of sedatives and analgesics; it also permits interactions between patients and their families. A common practice is to inject a bolus of midazolam (2–4 mg) or lorazepam (2–4 mg) before withdrawal, followed by a bolus of 5–10 mg of morphine and continuous infusion of morphine (50%

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

Some recommend terminal weaning because patients do not develop upper airway obstruction and the distress caused by secretions or stridor; however, terminal weaning can prolong the dying process and not allow a patient's family to be with the patient unencumbered by an endotracheal tube. To ensure comfort for conscious or semiconscious patients before withdrawal of the ventilator, neuromuscular blocking agents should be terminated and sedatives and analgesics administered. Removing the neuromuscular blocking agents permits patients to show discomfort, facilitating the titration of sedatives and analgesics; it also permits interactions between patients and their families. A common practice is to inject a bolus of midazolam (2–4 mg) or lorazepam (2–4 mg) before withdrawal, followed by a bolus of 5–10 mg of morphine and continuous infusion of morphine (50% of the bolus dose per hour) during weaning. In patients who have significant upper airway secretions, IV scopolamine at a rate of 100 µg/h can be administered. Additional boluses of morphine or increases in the infusion rate should be administered for respiratory distress or signs of pain. Higher doses will be needed for patients already receiving sedatives and opioids.

The median time to death after stopping of the ventilator is 1 h. However, up to 10% of patients unexpectedly survive for 1 day or more after mechanical ventilation is stopped. Women and older patients tend to survive longer after extubation. Families need to be reassured about both the continuations of treatments for common symptoms, such as dyspnea and agitation, after withdrawal of ventilatory support and the uncertainty of length of survival after withdrawal of ventilatory support.

FUTILE CARE

Beginning in the late 1980s, some commentators argued that physicians could terminate futile treatments demanded by the families of terminally ill patients. Although no objective definition or standard of futility exists, several categories have been proposed. Physiologic futility means that an intervention will have no physiologic effect. Some have defined qualitative futility as applying to procedures that "fail to end a patient's total dependence on intensive medical care." Quantitative futility occurs "when physicians conclude (through personal experience, experiences shared with colleagues, or consideration of reported empiric data) that in the last 100 cases, a medical treatment has been useless." The term conceals subjective value judgments about when a treatment is "not beneficial." Deciding whether a treatment that obtains an additional 6 weeks of life or a 1% survival advantage confers benefit depends on patients' preferences and goals. Furthermore, physicians' predictions of when treatments are futile deviate markedly from the quantitative definition. When residents thought CPR was quantitatively futile, more than one in five patients had a >10% chance of survival to hospital discharge. Most studies that purport to guide determinations of futility are based on insufficient data and therefore cannot provide statistical confidence for clinical decision-making. Quantitative futility rarely applies in ICU settings.

Many commentators reject using futility as a criterion for withdrawing care, preferring instead to consider futility situations as ones that represent conflict that calls for careful negotiation between families and health care providers. The American Medical Association and other professional societies have developed process-based approaches to resolving cases clinicians feel are futile. These process-based measures mainly suggest involving consultants and/or ethics committees when there are seemingly irresolvable differences. Some hospitals have enacted "unilateral do-not-resuscitate" policies to allow clinicians to provide a do-not-resuscitate order in cases in which consensus cannot be reached with families and medical opinion is that resuscitation would be futile if attempted. This type of a policy is not a replacement for careful and patient communication and negotiation but recognizes that agreement cannot always be reached.

In 1999, Texas enacted the so-called Futile Care Act. Other states, such as Virginia, Maryland, and California, have also enacted such laws that provide physicians a "safe harbor" from liability if they refuse a patient's or family's request for life-sustaining interventions. For instance, in Texas, when a disagreement about terminating interventions between the medical team and the family has not been resolved by an ethics consultation, the physician is tasked with trying to facilitate transfer of the patient to an institution willing to provide treatment. If this fails after 10 days, the hospital and physician may unilaterally withdraw treatments determined to be futile. The family may appeal to a state court. Early data suggest that the law increases futility consultations for the ethics committee and that, although most families concur with withdrawal, ~10–15% of families refuse to withdraw treatment. As of 2007, there had been 974 ethics committee consultations on medical futility cases and 65 in which committees ruled against families and gave notice that treatment would be terminated. In 2007, a survey of Texas hospitals showed that 30% of hospitals had used the futility law in 213 adult cases and 42 pediatric cases. Treatment was withdrawn for 27 of those patients, and the remainder were transferred to other facilities or died while awaiting transfer.

EUTHANASIA AND PHYSICIAN-ASSISTED SUICIDE

Euthanasia and PAS are defined in **Table 12-8**. Terminating life-sustaining care and providing opioid medications to manage symptoms such as pain or dyspnea have long been considered ethical by the medical profession and legal by courts and should not be conflated with euthanasia or PAS.

TABLE 12-8

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TERM	DEFINITION	LEGAL STATUS

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

of dyspnea have long been considered ethical by the medical profession and legal by courts and should not be conflated with euthanasia or PAS.

TABLE 12-8

Definitions of Physician-Assisted Suicide and Euthanasia

TERM	DEFINITION	LEGAL STATUS
Voluntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death with the patient's informed consent	Netherlands, Belgium, Luxembourg, Canada, Colombia, Spain, Western Australia, New Zealand
Involuntary active euthanasia	Intentionally administering medications or other interventions to cause the patient's death when the patient was competent to consent but did not—e.g., the patient may not have been asked	Nowhere
Passive euthanasia	Withholding or withdrawing life-sustaining medical treatments from a patient to let him or her die (terminating life-sustaining treatments)	Everywhere
Physician-assisted suicide	A physician provides medications or other interventions to a patient with the understanding that the patient can use them to commit suicide	Netherlands, Belgium, Luxembourg, Canada, Colombia, Germany, Switzerland, Oregon, Washington, Montana, Vermont, California, Colorado, District of Columbia, Hawaii, Maine, New Jersey, New Mexico

Legal Aspects

Euthanasia and PAS are legal in the Netherlands, Belgium, Luxembourg, Colombia, Canada, Spain, Western Australia, and New Zealand. Euthanasia was legalized in the Northern Territory of Australia in 1996, but that legislation was repealed 9 months later in 1997. Under certain conditions, a layperson in Switzerland or Germany can legally elect assisted suicide. In the United States, PAS is legal in Washington, D.C., and 10 states: Oregon, Washington State, Montana, Vermont, California, Colorado, Hawaii, Maine, New Jersey, and New Mexico. No state in the United States has legalized euthanasia. In the United States, multiple criteria must be met for PAS: the patient must have a terminal condition of <6 months and must be determined eligible through a process that includes a 15-day waiting period. In 2009, the state supreme court of Montana ruled that state law permits PAS for terminally ill patients. Many other countries, such as Portugal, are actively debating the legalization of euthanasia and/or PAS.

Practices

Fewer than 10–20% of terminally ill patients actually consider euthanasia and/or PAS for themselves. Use of euthanasia and PAS is increasing but remains relatively rare. In all countries, even the Netherlands and Belgium where these practices have been tolerated and legal for many years, <5% of death occur by euthanasia or PAS. As of the most recent data, 4.7% of all deaths were by euthanasia or PAS in the Netherlands (2015) and 4.6% in Belgium (2013). Just 0.50% of all deaths in Oregon in 2019 (188 of 37,397 deaths) and 0.36% of all deaths in Washington State in 2018 (203 of 56,913 deaths) were reported to be by PAS, although these may be underestimates since the cause of some deaths of patients who received medications could not be verified.

In Belgium, the Netherlands, Oregon, and Washington, >70% of patients utilizing these interventions are dying of cancer; <10% of deaths by euthanasia or PAS involve patients with AIDS or amyotrophic lateral sclerosis. While the numbers are small, in the Netherlands, the numbers of euthanasia or PAS cases in patients with psychiatric disorders, dementia, and the accumulation of health issues are increasing.

Pain is not the primary motivator for patients' requests for or interest in euthanasia and/or PAS. Among the first patients to receive PAS in Oregon, only 1 of the 15 patients had inadequate pain control, compared with 15 of the 43 patients in a control group who experienced inadequate pain relief. About 33% of patients in Oregon seeking PAS currently cite pain or fear of pain as their main reason for doing so. Conversely, depression and hopelessness are strongly associated with patient interest in euthanasia and PAS. Concerns about loss of dignity or autonomy or being a burden on

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

depressed terminally ill cancer patients were four times more likely to request euthanasia and confirmed that uncontrolled pain was not associated with greater interest in euthanasia.

hopelessness are strongly associated with patient interest in euthanasia and PAS. Concerns about loss of dignity or autonomy or being a burden on family members appear to be more important factors motivating a desire for euthanasia or PAS. Losing autonomy (87% Oregon [OR], 85% Washington [WA]), not being able to enjoy activities (90% OR, 84% WA), and fear of losing dignity (72% OR, 69% WA) are the most-cited end-of-life concerns in both states. A high percentage of patients seeking PAS note being a burden on family (59% OR, 51% WA). A study from the Netherlands showed that depressed terminally ill cancer patients were four times more likely to request euthanasia and confirmed that uncontrolled pain was not associated with greater interest in euthanasia.

Euthanasia and PAS are no guarantee of a painless, quick death. Data from the Netherlands indicate that in as many as 20% of euthanasia and PAS cases technical and other problems arose, including patients waking from coma, not becoming comatose, regurgitating medications, and experiencing a prolonged time to death. Data from Oregon between 1998 and 2017 and Washington between 2009 and 2017 indicate that of patients who received PAS prescriptions, 81% died at home and prescribers were present in 9.7% of cases. The time between drug intake and coma ranged from 1 min to 11 h, and the time from drug intake to death ranged from 1 min to 104 h. The median time from ingestion to coma was 5 min and from ingestion to death was 25 min. In Oregon between 1998 and 2015, 53% of patients had no complications, 44% of patients had no data on complications, and 2.4% of patients had regurgitation after taking the prescribed medicine as the only complication. In addition, six patients awakened. In Washington State between 2014 and 2015, 1.4% of patients had regurgitation, one patient had a seizure, and the reported range of time to death extended to 30 h. In the Netherlands, problems were significantly more common in PAS, sometimes requiring the physician to intervene and provide euthanasia.

Regardless of whether they practice in a setting where euthanasia is legal or not, many physicians over the course of their careers will receive a patient request for euthanasia or PAS. In the United States, 18% of physicians have received a request for PAS and 11% have received a request for euthanasia. Three percent complied with a request for PAS, while 5% complied with a request for euthanasia. In the Netherlands, where the practices are legal, 77% of physicians have received a request for PAS or euthanasia and 60% have performed these interventions.

Competency in dealing with such a request is crucial. Although challenging, the request can also provide a chance to address intense suffering. After receiving a request for euthanasia and/or PAS, health care providers should carefully clarify the request with empathic, open-ended questions to help elucidate the underlying cause for the request, such as, “What makes you want to consider this option?” Endorsing either moral opposition or moral support for the act tends to be counterproductive, giving an impression of being judgmental or of endorsing the idea that the patient’s life is worthless. Health care providers must reassure the patient of continued care and commitment. The patient should be educated about alternative, less laden options, such as symptom management and withdrawing any unwanted treatments, and the reality of euthanasia and/or PAS, since the patient may have misconceptions about their effectiveness as well as the legal implications of the choice. Depression, hopelessness, and other symptoms of psychological distress, as well as physical suffering and economic burdens, are likely factors motivating the request, and such factors should be assessed and treated aggressively. After these interventions and clarification of options, most patients proceed with another approach, declining life-sustaining interventions, possibly including refusal of nutrition and hydration.

CARE DURING THE LAST HOURS

Most laypersons have limited experiences with the actual dying process and death. They frequently do not know what to expect of the final hours and afterward. The family and other caregivers must be prepared, especially if the plan is for the patient to die at home.

Patients in the last days of life typically experience extreme weakness and fatigue and become bedbound; this can lead to pressure sores. The issue of turning patients who are near the end of life, however, must be balanced against the potential discomfort that movement may cause. Patients stop eating and drinking with drying of mucosal membranes and dysphagia. Careful attention to oral swabbing, lubricants for lips, and use of artificial tears can provide a form of care to substitute for attempts at feeding the patient. With loss of the gag reflex and dysphagia, patients may also experience accumulation of oral secretions, producing noises during respiration sometimes called “the death rattle.” Scopolamine can reduce the secretions. Patients also experience changes in respiration with periods of apnea or Cheyne-Stokes breathing. Decreased intravascular volume and cardiac output cause tachycardia, hypotension, peripheral coolness, and livedo reticularis (skin mottling). Patients can have urinary and, less frequently, fecal incontinence. Changes in consciousness and neurologic function generally lead to two different paths to death.

Each of these terminal changes can cause patients and families distress, requiring reassurance and targeted interventions **(Table 12-9)**. Informing families that these changes might occur and providing them with an information sheet can help preempt problems and minimize distress. Understanding that patients stop eating because they are dying, not dying because they have stopped eating, can reduce family and caregiver anxiety. Similarly, informing the family and caregivers that the “death rattle” may occur and that it is not indicative of suffocation, choking, or pain can reduce their worry from the breathing sounds.

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

CHANGES IN		FAMILY'S POSSIBLE	

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

Each of these terminal changes can cause patients and families distress, requiring reassurance and targeted interventions (Table 12-9). Informing their worry from the breathing sounds.

TABLE 12-9

Managing Changes in the Patient's Condition during the Final Days and Hours

CHANGES IN THE PATIENT'S CONDITION	POTENTIAL COMPLICATION	FAMILY'S POSSIBLE REACTION AND CONCERN	ADVICE AND INTERVENTION
Profound fatigue	Bedbound with development of pressure ulcers that are prone to infection, malodor, and pain, and joint pain	Patient is lazy and giving up.	Reassure family and caregivers that terminal fatigue will not respond to interventions and should not be resisted. Use an air mattress if necessary.
Anorexia	None	Patient is giving up; patient will suffer from hunger and will starve to death.	Reassure family and caregivers that the patient is not eating because he or she is dying; not eating at the end of life does not cause suffering or death. Forced feeding, whether oral, parenteral, or enteral, does not reduce symptoms or prolong life.
Dehydration	Dry mucosal membranes (see below)	Patient will suffer from thirst and die of dehydration.	Reassure family and caregivers that dehydration at the end of life does not cause suffering because patients lose consciousness before any symptom distress. Intravenous hydration can worsen symptoms of dyspnea by pulmonary edema and peripheral edema as well as prolong the dying process.
Dysphagia	Inability to swallow oral medications needed for palliative care		Do not force oral intake. Discontinue unnecessary medications that may have been continued, including antibiotics, diuretics, antidepressants, and laxatives. If swallowing pills is difficult, convert essential medications (analgesics, antiemetics, anxiolytics, and psychotropics) to oral solutions, buccal, sublingual, or rectal administration.
"Death rattle"—noisy breathing		Patient is choking and suffocating.	Reassure the family and caregivers that this is caused by secretions in the oropharynx and the patient is not choking. Reduce secretions with scopolamine (0.2–0.4 mg SC q4h or 1–3 patches q3d). Reposition patient to permit drainage of secretions. Do not suction. Suction can cause patient and family discomfort and is usually ineffective.
Apnea, Cheyne-Stokes respirations, dyspnea		Patient is suffocating.	Reassure family and caregivers that unconscious patients do not experience suffocation or air hunger. Apneic episodes are frequently a premorbid change. Opioids or anxiolytics may be used for dyspnea. Oxygen is unlikely to relieve dyspneic symptoms and may prolong the dying process.
Urinary or	Skin breakdown if days until death	Patient is dirty	Remind family and caregivers to use universal precautions
			urine output.
Agitation or	Dry/right to wound	Patient is in horrible pain	Reassure family and caregivers that agitation and delirium do

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

respirations			Apneic episodes are frequently a premorbid change
Urinary or fecal incontinence	Skin breakdown if days until death Potential transmission of infectious agents to caregivers	Patient is dirty, malodorous, and physically repellent.	Remind family and caregivers to use universal precautions. Frequent changes of bedclothes and bedding. Use diapers, urinary catheter, or rectal tube if diarrhea or high urine output.
Agitation or delirium	Day/night reversal Hurt self or caregivers	Patient is in horrible pain and going to have a horrible death.	Reassure family and caregivers that agitation and delirium do not necessarily connote physical pain. Depending on the prognosis and goals of treatment, consider evaluating for causes of delirium and modifying medications. Manage symptoms with haloperidol, chlorpromazine, diazepam, or midazolam.
Dry mucosal membranes	Cracked lips, mouth sores, and candidiasis can also cause pain. Odor	Patient may be malodorous, physically repellent.	Use baking soda mouthwash or saliva preparation q15–30 min. Use topical nystatin for candidiasis. Coat lips and nasal mucosa with petroleum jelly q60–90 min. Use ophthalmic lubricants q4h or artificial tears q30 min.

Families and caregivers may also feel guilty about stopping treatments, fearing that they are “killing” the patient. This may lead to demands for interventions, such as feeding tubes, that may be ineffective. In such cases, the physician should remind the family and caregivers about the inevitability of events and the palliative goals. Interventions may prolong the dying process and cause discomfort. Physicians also should emphasize that withholding treatments is both legal and ethical and that the family members are not the cause of the patient’s death. This reassurance may have to be provided multiple times.

Hearing and touch are said to be the last senses to stop functioning. Whether this is the case or not, families and caregivers can be encouraged to communicate with the dying patient. Encouraging them to talk directly to the patient, even if he or she is unconscious, and hold the patient’s hand or demonstrate affection in other ways can be an effective way to channel their urge “to do something” for the patient.

When the plan is for the patient to die at home, the physician must inform the family and caregivers how to determine that the patient has died. The cardinal signs are cessation of cardiac function and respiration; the pupils become fixed; the body becomes cool; muscles relax; and incontinence may occur. Remind the family and caregivers that the eyes may remain open even after the patient has died.

The physician should establish a plan for who the family or caregivers will contact when the patient is dying or has died. Without a plan, family members may panic and call 911, unleashing a cascade of unwanted events, from arrival of emergency personnel and resuscitation to hospital admission. The family and caregivers should be instructed to contact the hospice (if one is involved), the covering physician, or the on-call member of the palliative care team. They should also be told that the medical examiner need not be called unless the state requires it for all deaths. Unless foul play is suspected, the health care team need not contact the medical examiner either.

Just after the patient dies, even the best-prepared family may experience shock and loss and be emotionally distraught. They need time to assimilate the event and be comforted. Health care providers are likely to find it meaningful to write a bereavement card or letter to the family. The purpose is to communicate about the patient, perhaps emphasizing the patient’s virtues and the honor it was to care for the patient, and to express concern for the family’s hardship. Some physicians attend the funerals of their patients. Although this is beyond any medical obligation, the presence of the physician can be a source of support to the grieving family and provides an opportunity for closure for the physician.

Death of a spouse is a strong predictor of poor health, and even mortality, for the surviving spouse. It may be important to alert the spouse’s physician about the death so that he or she is aware of symptoms that might require professional attention.

FURTHER READING

Emanuel E et al: Attitudes and practices of euthanasia and physician-assisted suicide in the United States, Canada, and Europe. JAMA 316:79, 2016. [PubMed: 27380345]

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Kelley AS et al: Hospice enrollment saves money for Medicare and improves care quality across a number of different lengths of stay. Health Aff 32:552, 2012.

[PubMed: 27380345]

Kelley AS, Meier DE: Palliative care—A shifting paradigm. N Engl J Med 363:781, 2010. [PubMed: 20818881]

Kelley AS et al: Hospice enrollment saves money for Medicare and improves care quality across a number of different lengths-of-stay. Health Aff 32:552, 2012.

Kelley AS et al: Palliative care for the seriously ill. N Engl J Med 373:747, 2015. [PubMed: 26287850]

Mack JW et al: Associations between end-of-life discussion characteristics and care received near death: A prospective cohort study. J Clin Oncol 30:4387, 2012. [PubMed: 23150700]

Murray SA et al: Illness trajectories and palliative care. BMJ 330:1007, 2005. [PubMed: 15860828]

Neuman P et al: Medicare per capita spending by age and service: New data highlights oldest beneficiaries. Health Aff (Millwood) 34:335, 2015. [PubMed: 25588646]

Nicholas LH et al: Regional variation in the association between advance directives and end-of-life Medicare expenditures. JAMA 306:1447, 2011. [PubMed: 21972306]

Ornstein KA et al: Evaluation of racial disparities in hospice use and end-of-life treatment intensity in the REGARDS cohort. JAMA Netw Open 3(8):e2014639, 2020. [PubMed: 32833020]

Quinn KL et al: Association of receipt of palliative care interventions with health care use, quality of life, and symptom burden among adults with chronic noncancer illness: A systematic review and meta-analysis. JAMA 324:1439, 2020. [PubMed: 33048152]

Teno JM et al: Change in end-of-life care for medicare beneficiaries: Site of death, place of care, and health transitions in 2000, 2005, and 2009. JAMA 309:470, 2013. [PubMed: 23385273]

Teno JM et al: Site of death, place of care, and health care transitions among US Medicare beneficiaries, 2000-2015. JAMA 320:264, 2018. [PubMed: 29946682]

Van Den Beuken-VanEverdingen MH et al: Update on prevalence of pain in patients with cancer: Systematic review and meta-analysis. J Pain Symptom Manage 51:1070, 2016. [PubMed: 27112310]

WEBSITES

American Academy of Hospice and Palliative Medicine: *www.aahpm.org*

Center to Advance Palliative Care: *http://www.capc.org*

Education in Palliative and End of Life Care (EPEC): *http://www.epec.net*

Family Caregiver Alliance: *http://www.caregiver.org*

National Hospice and Palliative Care Organization (including state-specific advance directives): *http://www.nhpco.org*

NCCN: The National Comprehensive Cancer Network palliative care guidelines: *http://www.nccn.org*

Our Care Wishes Advance Care Planning Tool: *https://www.ourcarewishes.org*

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Downloaded 2022-1-3 5:42 P Your IP is 141.216.78.40
Chapter 12: Palliative and End-of-Life Care, Ezekiel J. Emanuel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 13: Pain: Pathophysiology and Management

James P. Rathmell; Howard L. Fields

INTRODUCTION

The province of medicine is to preserve and restore health and to relieve suffering. Understanding pain is essential to both of these goals. Because pain is universally understood as a signal of disease, it is the most common symptom that brings a patient to a physician's attention. The function of the pain sensory system is to protect the body and maintain homeostasis. It does this by detecting, localizing, and identifying potential or actual tissue-damaging processes. Because different diseases produce characteristic patterns of tissue damage, the quality, time course, and location of a patient's pain lend important diagnostic clues. It is the physician's responsibility to assess each patient promptly for any remediable cause underlying the pain and to provide rapid and effective pain relief whenever possible.

THE PAIN SENSORY SYSTEM

Pain is an unpleasant sensation localized to a part of the body. It is often described in terms of a penetrating or tissue-destructive process (e.g., stabbing, burning, twisting, tearing, squeezing) and/or of a bodily or emotional reaction (e.g., terrifying, nauseating, sickening). Furthermore, any pain of moderate or higher intensity is accompanied by anxiety and the urge to escape or terminate the feeling. These properties illustrate the duality of pain: it is both sensation and emotion. When it is acute, pain is characteristically associated with behavioral arousal and a stress response consisting of increased blood pressure, heart rate, pupil diameter, and plasma cortisol levels. In addition, local muscle contraction (e.g., limb flexion, abdominal wall rigidity) is often present.

PERIPHERAL MECHANISMS

The Primary Afferent Nociceptor

A peripheral nerve consists of the axons of three different types of neurons: primary sensory afferents, motor neurons, and sympathetic postganglionic neurons **(Fig. 13-1)**. The cell bodies of primary sensory afferents are located in the dorsal root ganglia within the vertebral foramina. The primary afferent axon has two branches: one projects centrally into the spinal cord and the other projects peripherally to innervate tissues. Primary afferents are classified by their diameter, degree of myelination, and conduction velocity. The largest diameter afferent fibers, A-beta (Aβ), respond maximally to light touch and/or moving stimuli; they are present primarily in nerves that innervate the skin. In normal individuals, the activity of these fibers does not produce pain. There are two other classes of primary afferent nerve fibers: the small diameter myelinated A-delta (Aδ) and the unmyelinated (C) axons (Fig. 13-1). These fibers are present in nerves to the skin and to deep somatic and visceral structures. Some tissues, such as the cornea, are innervated only by Aδ and C fiber afferents. Most Aδ and C fiber afferents respond maximally to intense (painful) stimuli and produce the subjective experience of pain when they are activated; this defines them as *primary afferent nociceptors* (*pain receptors*). The ability to detect painful stimuli is completely abolished when conduction in Aδ and C fiber axons is blocked.

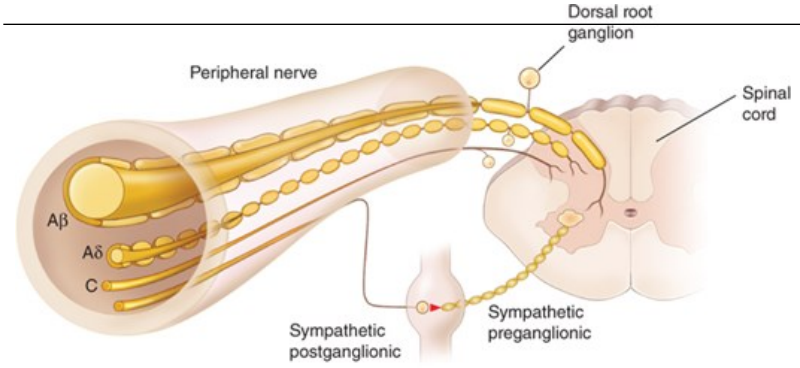
FIGURE 13-1

Components of a typical cutaneous nerve. There are two distinct functional categories of axons: primary afferents with cell bodies in the dorsal root ganglion and sympathetic postganglionic fibers with cell bodies in the sympathetic ganglion. Primary afferents include those with large-diameter myelinated (Aβ), small-diameter myelinated (Aδ), and unmyelinated (C) axons. All sympathetic postganglionic fibers are unmyelinated.



Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Individual primary afferent nociceptors can respond to several different types of noxious stimuli. For example, most nociceptors respond to heat; intense cold; intense mechanical distortion, such as a pinch; changes in pH, particularly an acidic environment; and application of chemical irritants including adenosine triphosphate (ATP), serotonin, bradykinin (BK), and histamine. The transient receptor potential cation channel subfamily V member 1 (TrpV1), also known as the vanilloid receptor, mediates perception of some noxious stimuli, especially heat sensations, by nociceptive neurons; it is activated by heat, acidic pH, endogenous mediators, and capsaicin, a component of hot chili peppers.

Sensitization

When intense, repeated, or prolonged stimuli are applied to damaged or inflamed tissues, the threshold for activating primary afferent nociceptors is lowered, and the frequency of firing is higher for all stimulus intensities. Inflammatory mediators such as BK, nerve-growth factor, some prostaglandins (PGs), and leukotrienes contribute to this process, which is called *sensitization*. Sensitization occurs at the level of the peripheral nerve terminal (*peripheral sensitization*) as well as at the level of the dorsal horn of the spinal cord (*central sensitization*). Peripheral sensitization occurs in damaged or inflamed tissues, when inflammatory mediators activate intracellular signal transduction in nociceptors, prompting an increase in the production, transport, and membrane insertion of chemically gated and voltage-gated ion channels. These changes increase the excitability of nociceptor terminals and lower their threshold for activation by mechanical, thermal, and chemical stimuli. Central sensitization occurs when activity, generated by nociceptors during inflammation, enhances the excitability of nerve cells in the dorsal horn of the spinal cord. Following injury and resultant sensitization, normally innocuous stimuli can produce pain (termed *allodynia*). Sensitization is a clinically important process that contributes to tenderness, soreness, and *hyperalgesia* (increased pain intensity in response to the same noxious stimulus; e.g., pinprick causes severe pain). A striking example of sensitization is sunburned skin, in which severe pain can be produced by a gentle slap or a warm shower.

Sensitization is of particular importance for pain and tenderness in deep tissues. Viscera are normally relatively insensitive to noxious mechanical and thermal stimuli, although hollow viscera do generate significant discomfort when distended. In contrast, when affected by a disease process with an inflammatory component, deep structures such as joints or hollow viscera characteristically become exquisitely sensitive to mechanical stimulation.

A large proportion of Aδ and C fiber afferents innervating viscera are completely insensitive in normal noninjured, noninflamed tissue. That is, they cannot be activated by known mechanical or thermal stimuli and are not spontaneously active. However, in the presence of inflammatory mediators, these afferents become sensitive to mechanical stimuli. Such afferents have been termed *silent nociceptors*, and their characteristic properties may explain how, under pathologic conditions, the relatively insensitive deep structures can become the source of severe and debilitating pain and tenderness. Low pH, PGs, leukotrienes, and other inflammatory mediators such as BK play a significant role in sensitization.

Nociceptor-Induced Inflammation

Primary afferent nociceptors are not simply passive messengers of threats to tissue injury but also play an active role in tissue protection through a neuroeffector function. Most nociceptors contain polypeptide mediators, including substance P, calcitonin gene related peptide (CGRP), and cholecystokinin, that are released from their peripheral terminals when they are activated **(Fig. 13-2)**. Substance P is an 11-amino-acid peptide that is released in peripheral tissues from primary afferent nociceptors and has multiple biologic activities. It is a potent vasodilator, causes mast cell degranulation, is a chemoattractant for leukocytes, and increases the production and release of inflammatory mediators. Interestingly, depletion of substance P from joints reduces the severity of experimental arthritis.

FIGURE 13-2

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

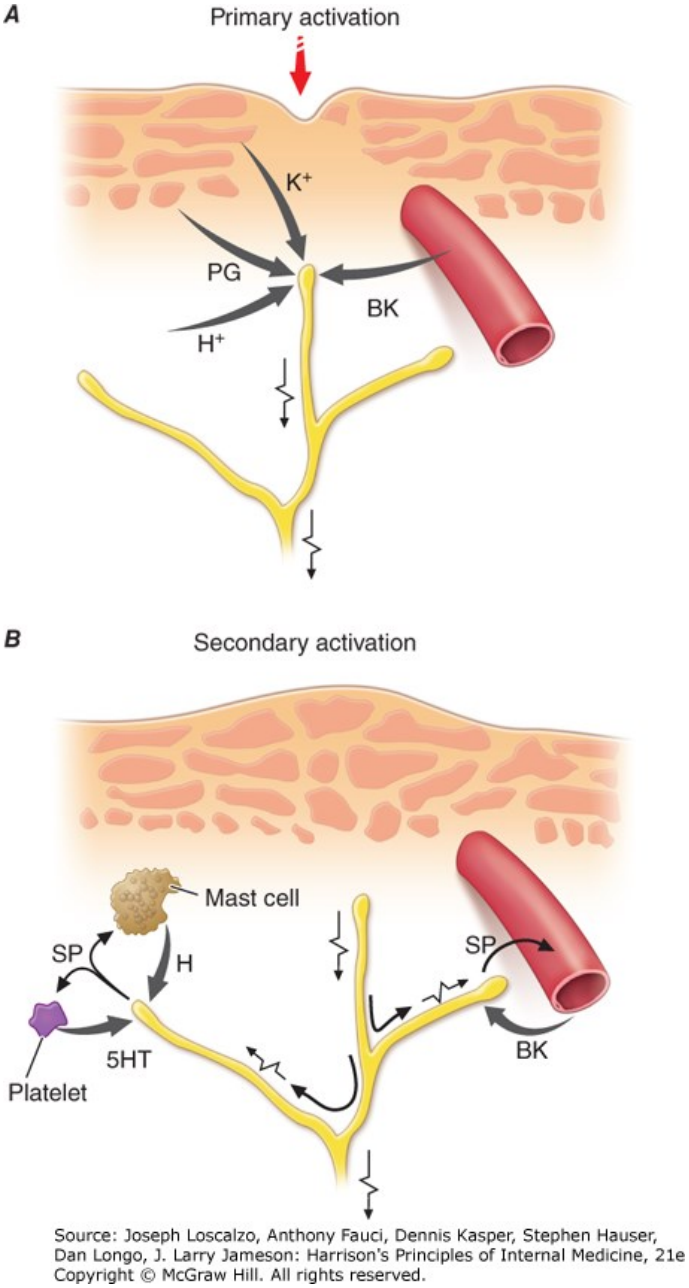
prostaglandins (PGs) and bradykinin (BK). PGs increase the sensitivity of the terminal to BK and other pain-producing substances. ***B***. Secondary

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

FIGURE 13-2

Events leading to activation, sensitization, and spread of sensitization of primary afferent nociceptor terminals. ***A***. Direct activation by intense pressure and consequent cell damage. Cell damage induces lower pH (H^+) and leads to release of potassium (K^+) and to synthesis of prostaglandins (PGs) and bradykinin (BK). PGs increase the sensitivity of the terminal to BK and other pain-producing substances. ***B***. Secondary activation. Impulses generated in the stimulated terminal propagate not only to the spinal cord but also into other terminal branches where they induce the release of peptides, including substance P (SP). Substance P causes vasodilation and neurogenic edema with further accumulation of BK. Substance P also causes the release of histamine (H) from mast cells and serotonin (5HT) from platelets.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

CENTRAL MECHANISMS

The Spinal Cord and Referred Pain

The axons of primary afferent nociceptors enter the spinal cord via the dorsal root. They terminate in the dorsal horn of the spinal gray matter (Fig.

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

release substance P and CGRP, which produce a slower and longer-lasting excitation of the dorsal horn neurons. The axon of each primary afferent contacts many spinal neurons, and each spinal neuron receives convergent inputs from many primary afferents.

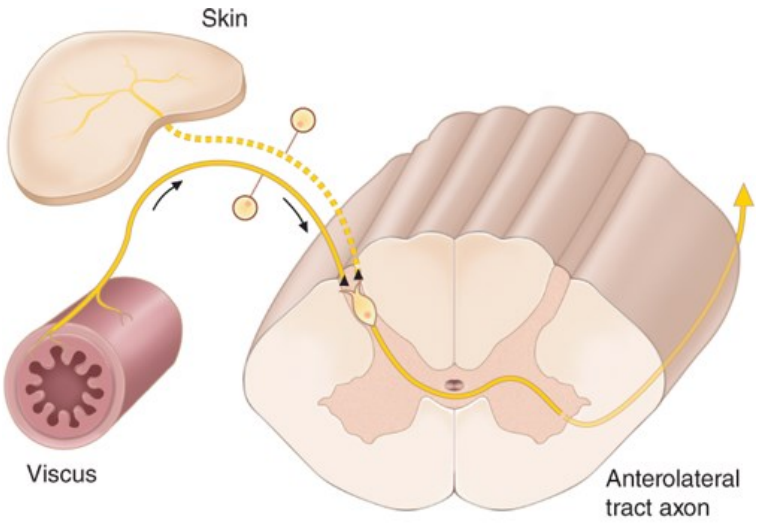
ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

The axons of primary afferent nociceptors enter the spinal cord via the dorsal root. They terminate in the dorsal horn of the spinal gray matter **(Fig. 13-3)**. The terminals of primary afferent axons contact spinal neurons that transmit the pain signal to brain sites involved in pain perception. When primary afferents are activated by noxious stimuli, they release neurotransmitters from their terminals that excite the spinal cord neurons. The major neurotransmitter released is glutamate, which rapidly excites the second-order dorsal horn neurons. Primary afferent nociceptor terminals also release substance P and CGRP, which produce a slower and longer-lasting excitation of the dorsal horn neurons. The axon of each primary afferent contacts many spinal neurons, and each spinal neuron receives convergent inputs from many primary afferents.

FIGURE 13-3

The convergence-projection hypothesis of referred pain. According to this hypothesis, visceral afferent nociceptors converge on the same pain-projection neurons as the afferents from the somatic structures in which the pain is perceived. The brain has no way of knowing the actual source of input and mistakenly "projects" the sensation to the somatic structure.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The convergence of sensory inputs to a single spinal pain-transmission neuron is of great importance because it underlies the phenomenon of referred pain. All spinal neurons that receive input from the viscera and deep musculoskeletal structures also receive input from the skin. The convergence patterns are determined by the spinal segment of the dorsal root ganglion that supplies the afferent innervation of a structure. For example, the afferents that supply the central diaphragm are derived from the third and fourth cervical dorsal root ganglia. Primary afferents with cell bodies in these same ganglia supply the skin of the shoulder and lower neck. Thus, sensory inputs from both the shoulder skin and the central diaphragm converge on pain-transmission neurons in the third and fourth cervical spinal segments. *Because of this convergence and the fact that the spinal neurons are most often activated by inputs from the skin, activity evoked in spinal neurons by input from deep structures is often mislocalized by the patient to a bodily location that roughly corresponds with the region of skin innervated by the same spinal segment.* Thus, inflammation near the central diaphragm is often reported as shoulder discomfort. This spatial displacement of pain sensation from the site of the injury that produces it is known as *referred pain*.

Ascending Pathways for Pain

A majority of spinal neurons contacted by primary afferent nociceptors send their axons to the contralateral thalamus. These axons form the contralateral spinothalamic tract, which lies in the anterolateral white matter of the spinal cord, the lateral edge of the medulla, and the lateral pons and midbrain. The spinothalamic pathway is crucial for pain sensation in humans. Interruption of this pathway produces permanent deficits in pain and temperature discrimination.

Spinothalamic tract axons ascend to several regions of the thalamus. There is tremendous divergence of the pain signal from these thalamic sites to several distinct areas of the cerebral cortex that subserve different aspects of the pain experience **(Fig. 13-4)**. One of the thalamic projections is to the somatosensory cortex. This projection mediates the sensory discriminative aspects of pain, i.e., its location, intensity, and quality. Other thalamic neurons project to cortical regions that are linked to emotional responses, such as the cingulate and insular cortex. These pathways to the frontal

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

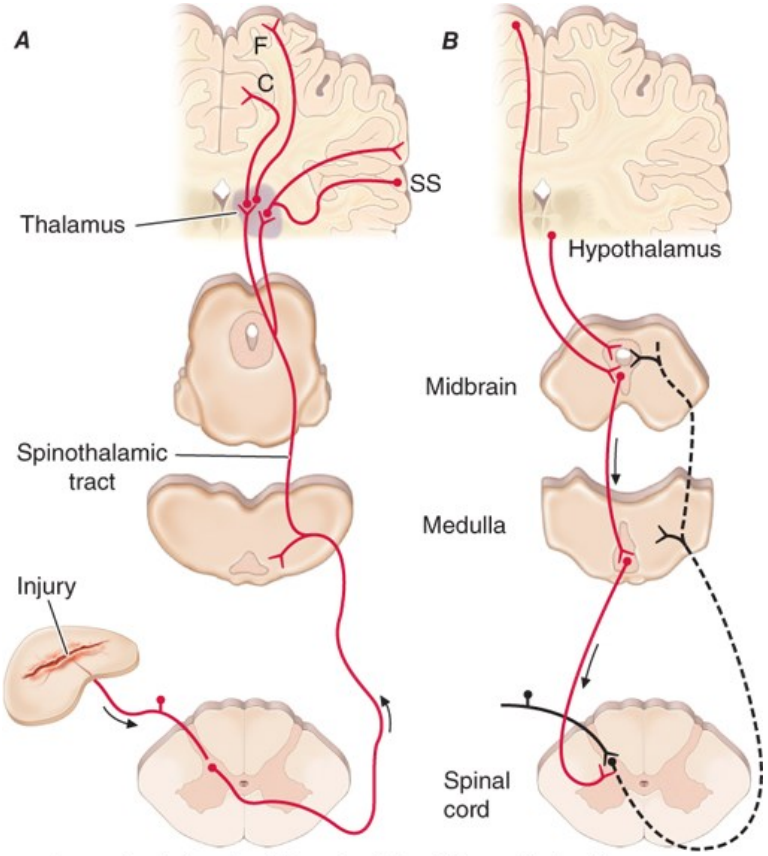
noxious stimuli as painful.

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

somatosensory cortex. This projection mediates the sensory discriminative aspects of pain, i.e., its location, intensity, and quality. Other thalamic neurons project to cortical regions that are linked to emotional responses, such as the cingulate and insular cortex. These pathways to the frontal cortex subserve the affective or unpleasant emotional dimension of pain. This affective dimension of pain produces suffering and exerts potent control of behavior. Because of this dimension, fear is a constant companion of pain. As a consequence, injury or surgical lesions to areas of the frontal cortex activated by painful stimuli can diminish the emotional impact of pain while largely preserving the individual's ability to recognize noxious stimuli as painful.

FIGURE 13-4

Pain-transmission and modulatory pathways. ***A***. Transmission system for nociceptive messages. Noxious stimuli activate the sensitive peripheral ending of the primary afferent nociceptor by the process of transduction. The message is then transmitted over the peripheral nerve to the spinal cord, where it synapses with cells of origin of the major ascending pain pathway, the spinothalamic tract. The message is relayed in the thalamus to the anterior cingulate (C), frontal insular (F), and somatosensory cortex (SS). ***B***. Pain-modulation network. Inputs from frontal cortex and hypothalamus activate cells in the midbrain that control spinal pain-transmission cells via cells in the medulla.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

PAIN MODULATION

The pain produced by injuries of similar magnitude is remarkably variable in different situations and in different individuals. For example, athletes have been known to sustain serious fractures with only minor pain, and Beecher's classic World War II survey revealed that many soldiers in battle were unbothered by injuries that would have produced agonizing pain in civilian patients. Furthermore, even the suggestion that a treatment will relieve pain can have a significant analgesic effect (the *placebo effect*). On the other hand, many patients find even minor injuries such as venipuncture frightening and unbearable, and the expectation of pain can induce pain even without a noxious stimulus. The suggestion that pain will worsen following administration of an inert substance can increase its perceived intensity (the *nocebo effect*).

The powerful effect of expectation and other psychological variables on the perceived intensity of pain is explained by brain circuits that modulate the activity of the pain-transmission pathways. One of these circuits has links to the hypothalamus, midbrain, and medulla, and it selectively controls

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

medications **(Fig. 13-5)**. Furthermore, each of the component structures of the pathway contains opioid receptors and is sensitive to the direct application of opioid drugs. In animals, lesions of this descending modulatory system reduce the analgesic effect of systemically administered opioids

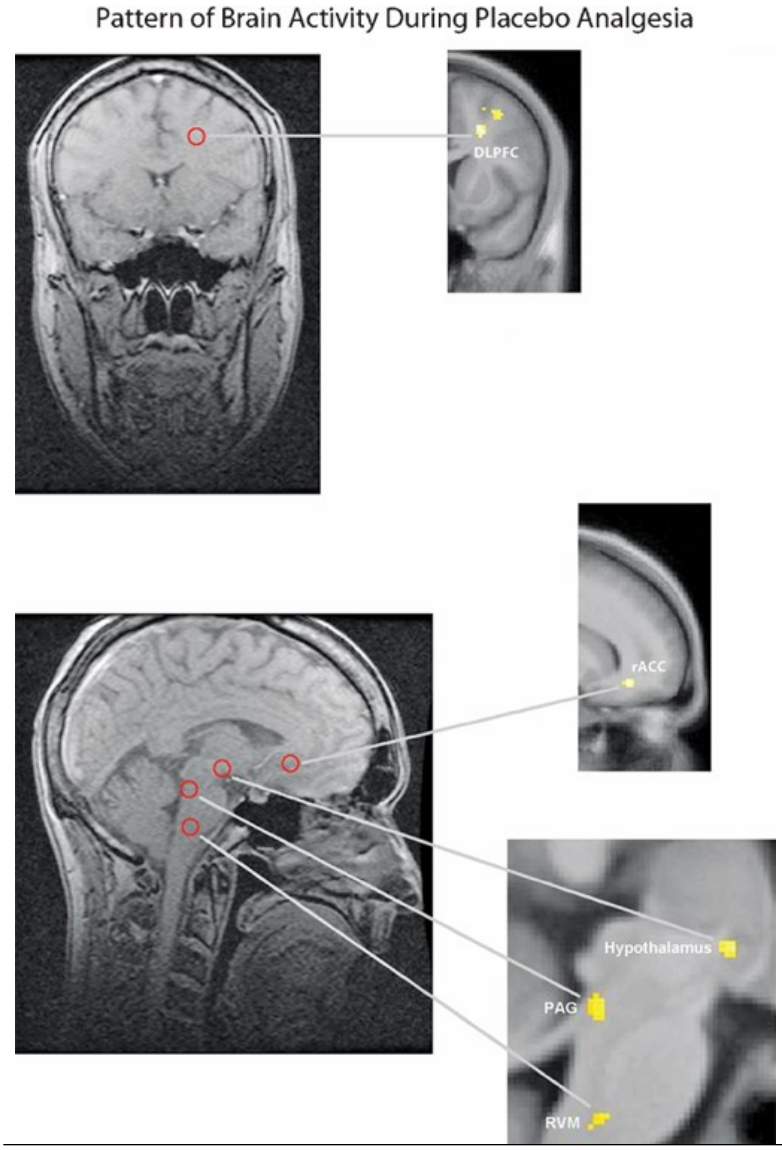
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

The powerful effect of expectation and other psychological variables on the perceived intensity of pain is explained by brain circuits that modulate the activity of the pain-transmission pathways. One of these circuits has links to the hypothalamus, midbrain, and medulla, and it selectively controls spinal pain-transmission neurons through a descending pathway (Fig. 13-4).

Human brain-imaging studies have implicated this pain-modulating circuit in the pain-relieving effect of attention, suggestion, and opioid analgesic medications **(Fig. 13-5)**. Furthermore, each of the component structures of the pathway contains opioid receptors and is sensitive to the direct application of opioid drugs. In animals, lesions of this descending modulatory system reduce the analgesic effect of systemically administered opioids such as morphine. Along with the opioid receptor, the component nuclei of this pain-modulating circuit contain endogenous opioid peptides such as the enkephalins and β-endorphin.

FIGURE 13-5

Functional magnetic resonance imaging (fMRI) demonstrates placebo-enhanced brain activity in anatomic regions correlating with the opioidergic descending pain control system. *Top panel:* Frontal fMRI image shows placebo-enhanced brain activity in the dorsal lateral prefrontal cortex (DLPFC). *Bottom panel:* Sagittal fMRI images show placebo-enhanced responses in the rostral anterior cingulate cortex (rACC), the rostral ventral medullae (RVM), the periaqueductal gray (PAG) area, and the hypothalamus. The placebo-enhanced activity in all areas was reduced by naloxone, demonstrating the link between the descending opioidergic system and the placebo analgesic response. *(Adapted with permission from F Eippert, U Bingel, ED Schoell et al: Activation of the opioidergic descending pain control system underlies placebo analgesia. Neuron 63(4):533–543, 2009.)*



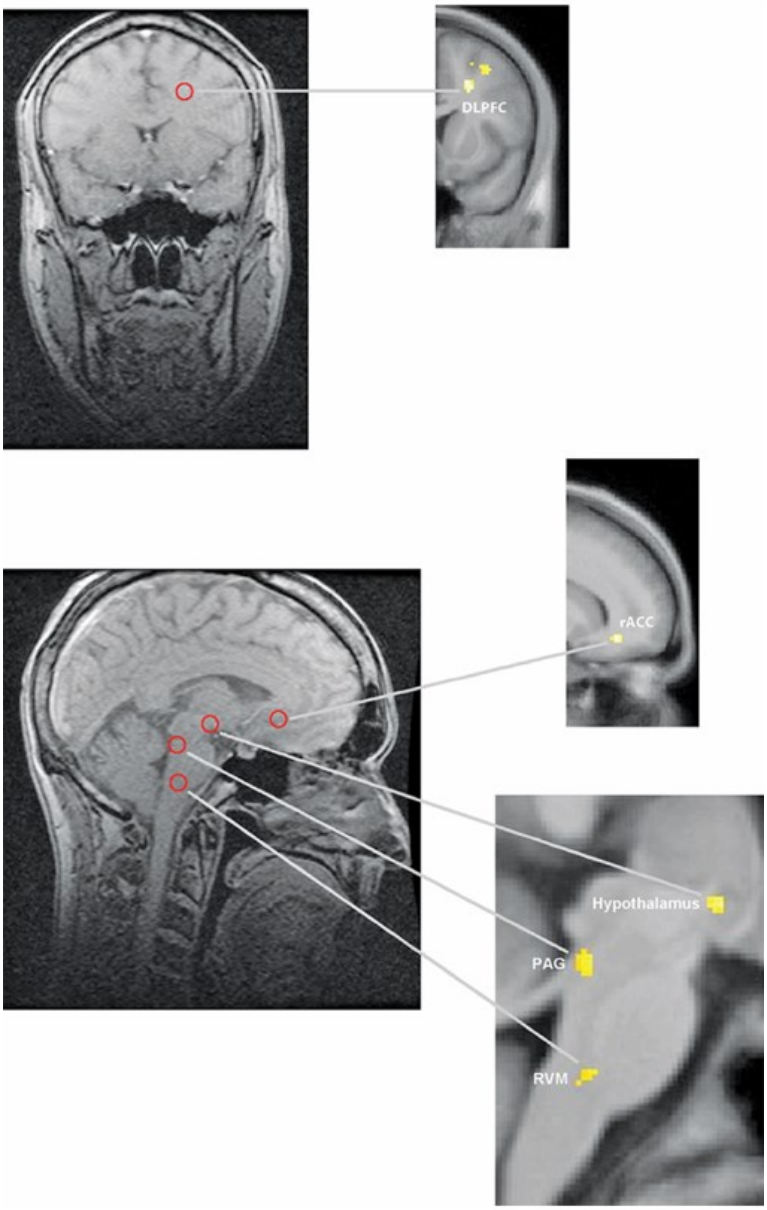
Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw-Hill. All rights reserved.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Pattern of Brain Activity During Placebo Analgesia



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The most reliable way to activate this endogenous opioid-mediated modulating system is by suggestion of pain relief or by intense emotion directed away from the pain-causing injury (e.g., during severe threat or an athletic competition). In fact, pain-relieving endogenous opioids are released following surgical procedures and in patients given a placebo for pain relief.

Pain-modulating circuits can enhance as well as suppress pain. Both pain-inhibiting and pain-facilitating neurons in the medulla project to and control spinal pain-transmission neurons. Because pain-transmission neurons can be activated by modulatory neurons, it is theoretically possible to generate a pain signal with no peripheral noxious stimulus. In fact, human functional imaging studies have demonstrated increased activity in this circuit during migraine headaches. A central circuit that facilitates pain could account for the finding that pain can be induced by suggestion or enhanced by expectation and provides a framework for understanding how psychological factors can contribute to chronic pain.

NEUROPATHIC PAIN

Lesions of the peripheral or central nociceptive pathways typically result in a loss or impairment of pain sensation. Paradoxically, damage to or dysfunction of these pathways can also produce pain. For example, damage to peripheral nerves, as occurs in diabetic neuropathy, or to primary

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

standard treatments for pain.

ACCESS Medicine

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Lesions of the peripheral or central nociceptive pathways typically result in a loss or impairment of pain sensation. Paradoxically, damage to or dysfunction of these pathways can also produce pain. For example, damage to peripheral nerves, as occurs in diabetic neuropathy, or to primary afferents, as in herpes zoster infection, can result in pain that is referred to the body region innervated by the damaged nerves. Pain may also be produced by damage to the central nervous system (CNS), for example, in some patients following trauma or vascular injury to the spinal cord, brainstem, or thalamic areas that contain central nociceptive pathways. Such pains are termed *neuropathic* and are often severe and resistant to standard treatments for pain.

Neuropathic pain typically has an unusual burning, tingling, or electric shock-like quality and may occur spontaneously, without any stimulus, or be triggered by very light touch. These features are rare in other types of pain. On examination, a sensory deficit is characteristically co-extensive with the area of the patient's pain. *Hyperpathia*, a greatly exaggerated pain response to innocuous or mild nociceptive stimuli, especially when applied repeatedly, is also characteristic of neuropathic pain; patients often complain that the very lightest moving stimulus evokes exquisite pain (allodynia). In this regard, it is of clinical interest that a topical preparation of 5% lidocaine in patch form is effective for patients with postherpetic neuralgia who have prominent allodynia.

A variety of mechanisms contribute to neuropathic pain. As with sensitized primary afferent nociceptors, damaged primary afferents, including nociceptors, become highly sensitive to mechanical stimulation and may generate impulses in the absence of stimulation. Increased sensitivity and spontaneous activity are due, in part, to an increased density of sodium channels in the damaged nerve fiber. Damaged primary afferents may also develop sensitivity to norepinephrine. Interestingly, spinal cord pain-transmission neurons cut off from their normal input may also become spontaneously active. Thus, both central and peripheral nervous system hyperactivity contribute to neuropathic pain.

Sympathetically Maintained Pain

Patients with peripheral nerve injury occasionally develop spontaneous pain in or beyond the region innervated by the nerve. This pain is often described as having a burning quality. The pain typically begins after a delay of hours to days or even weeks and is accompanied by swelling of the extremity, periarticular bone loss, and arthritic changes in the distal joints. Early in the course of the condition, the pain may be relieved by a local anesthetic block of the sympathetic innervation to the affected extremity. Damaged primary afferent nociceptors acquire adrenergic sensitivity and can be activated by stimulation of the sympathetic outflow. This constellation of spontaneous pain and signs of sympathetic dysfunction following injury has been termed *complex regional pain syndrome* (CRPS). When this occurs after an identifiable nerve injury, it is termed CRPS type II (also known as posttraumatic neuralgia or, if severe, *causalgia*). When a similar clinical picture appears without obvious nerve injury, it is termed CRPS type I (also known as *reflex sympathetic dystrophy*). CRPS can be produced by a variety of injuries, including fractures of bone, soft tissue trauma, myocardial infarction, and stroke. CRPS type I typically resolves with symptomatic treatment; however, when it persists, detailed examination often reveals evidence of peripheral nerve injury. Although the pathophysiology of CRPS is poorly understood, the pain and the signs of inflammation, when acute, can be rapidly relieved by blocking the sympathetic nervous system. This implies that sympathetic activity can activate undamaged nociceptors when inflammation is present. Signs of sympathetic hyperactivity should be sought in patients with posttraumatic pain and inflammation and no other obvious explanation.

TREATMENT

Acute Pain

The ideal treatment for any pain is to remove the cause; thus, while treatment can be initiated immediately, efforts to establish the underlying etiology should always proceed as treatment begins. Sometimes, treating the underlying condition does not immediately relieve pain. Furthermore, some conditions are so painful that rapid and effective analgesia is essential (e.g., the postoperative state, burns, trauma, cancer, or sickle cell crisis). Analgesic medications are a first line of treatment in these cases, and all practitioners should be familiar with their use.

ASPIRIN, ACETAMINOPHEN, AND NONSTEROIDAL ANTI-INFLAMMATORY AGENTS (NSAIDS)

These drugs are considered together because they are used for similar problems and may have a similar mechanism of action **(Table 13-1)**. All these compounds inhibit cyclooxygenase (COX), and except for acetaminophen, all have anti-inflammatory actions, especially at higher dosages. They are particularly effective for mild to moderate headache and for pain of musculoskeletal origin.

Because they are effective for these common types of pain and are available without prescription, COX inhibitors are by far the most commonly used analgesics. They are absorbed well from the gastrointestinal tract and, with occasional use, have only minimal side effects. With chronic use, gastric irritation is a common side effect of aspirin and NSAIDs and is the problem that most frequently limits the dose that can be given. Gastric irritation is

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

nephrotoxicity is a significant problem for patients using these drugs on a chronic basis. Patients at risk for renal insufficiency, particularly those with significant contraction of their intravascular volume as occurs with chronic diuretic use or acute hypovolemia, should avoid NSAIDs. NSAIDs can

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

analgesics. They are absorbed well from the gastrointestinal tract and, with occasional use, have only minimal side effects. With chronic use, gastric irritation is a common side effect of aspirin and NSAIDs and is the problem that most frequently limits the dose that can be given. Gastric irritation is most severe with aspirin, which may cause erosion and ulceration of the gastric mucosa leading to bleeding or perforation. Because aspirin irreversibly acetylates platelet COX and thereby interferes with coagulation of the blood, gastrointestinal bleeding is a particular risk. Older age and history of gastrointestinal disease increase the risks of aspirin and NSAIDs. In addition to the well-known gastrointestinal toxicity of NSAIDs, nephrotoxicity is a significant problem for patients using these drugs on a chronic basis. Patients at risk for renal insufficiency, particularly those with significant contraction of their intravascular volume as occurs with chronic diuretic use or acute hypovolemia, should avoid NSAIDs. NSAIDs can also increase blood pressure in some individuals. Long-term treatment with NSAIDs requires regular blood pressure monitoring and treatment if necessary. Although toxic to the liver when taken in high doses, acetaminophen rarely produces gastric irritation and does not interfere with platelet function.

The introduction of parenteral forms of NSAIDs, ketorolac and diclofenac, extends the usefulness of this class of compounds in the management of acute severe pain. Both agents are sufficiently potent and rapid in onset to supplant opioids as first-line treatment for many patients with acute severe headache and musculoskeletal pain.

There are two major classes of COX: COX-1 is constitutively expressed, and COX-2 is induced in the inflammatory state. COX-2-selective drugs have similar analgesic potency and produce less gastric irritation than the nonselective COX inhibitors. The use of COX-2-selective drugs does not appear to lower the risk of nephrotoxicity compared to nonselective NSAIDs. On the other hand, COX-2-selective drugs offer a significant benefit in the management of acute postoperative pain because they do not affect blood coagulation. Nonselective COX inhibitors (especially aspirin) are usually contraindicated postoperatively because they impair platelet-mediated blood clotting and are thus associated with increased bleeding at the operative site. COX-2 inhibitors, including celecoxib (Celebrex), are associated with increased cardiovascular risk, including cardiovascular death, myocardial infarction, stroke, heart failure, or a thromboembolic event. It appears that this is a class effect of NSAIDs, excluding aspirin. These drugs are contraindicated in patients in the immediate period after coronary artery bypass surgery and should be used with caution in elderly patients and those with a history of or significant risk factors for cardiovascular disease.

OPIOID ANALGESICS

Opioids are the most potent pain-relieving drugs currently available. Of all analgesics, they have the broadest range of efficacy and provide the most reliable and effective treatment for rapid pain relief. Although side effects are common, most are reversible: nausea, vomiting, pruritus, sedation, and constipation are the most frequent and bothersome side effects. Respiratory depression is uncommon at standard analgesic doses but can be life-threatening. Opioid-related side effects can be reversed rapidly with the narcotic antagonist naloxone. Many physicians, nurses, and patients have a certain trepidation about using opioids that is based on a fear of initiating addiction in their patients. In fact, there is a very small chance of patients becoming addicted to narcotics as a result of their appropriate medical use. For chronic pain, particularly chronic noncancer pain, the risk of addiction in patients taking opioids on a chronic basis remains small, but the risk does appear to increase with dose escalation. The physician should not hesitate to use opioid analgesics in patients with acute severe pain. Table 13-1 lists the most commonly used opioid analgesics.

Opioids produce analgesia by actions in the CNS. They activate pain-inhibitory neurons and directly inhibit pain-transmission neurons. Most of the commercially available opioid analgesics act at the same opioid receptor (μ-receptor), differing mainly in potency, speed of onset, duration of action, and optimal route of administration. Some side effects are due to accumulation of nonopioid metabolites that are unique to individual drugs. One striking example of this is normeperidine, a metabolite of meperidine. At higher doses of meperidine, typically >1 g/d, accumulation of normeperidine can produce hyperexcitability and seizures that are not reversible with naloxone. Normeperidine accumulation is increased in patients with renal failure.

The most rapid pain relief is obtained by intravenous administration of opioids; relief with oral administration is significantly slower. Because of the potential for respiratory depression, patients with any form of respiratory compromise must be kept under close observation following opioid administration; an oxygen-saturation monitor may be useful, but only in a setting where the monitor is under constant surveillance. Opioid-induced respiratory depression is primarily manifest as a reduction in respiratory rate and is typically accompanied by sedation. A fall in oxygen saturation represents a critical level of respiratory depression and the need for immediate intervention to prevent life-threatening hypoxemia. Newer monitoring devices that incorporate capnography or pharyngeal air flow can detect apnea at the point of onset and should be used in hospitalized patients. Ventilatory assistance should be maintained until the opioid-induced respiratory depression has resolved. The opioid antagonist naloxone should be readily available whenever opioids are used at high doses or in patients with compromised pulmonary function. Opioid effects are dose-related, and there is great variability among patients in the doses that relieve pain and produce side effects. Synergistic respiratory depression is common when opioids are administered with other CNS depressants. Co-administration of benzodiazepines is particularly likely to produce

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

made by physicians in managing severe pain with opioids is to prescribe an inadequate dose. Because many patients are reluctant to complain, this practice leads to needless suffering. In the absence of sedation at the expected time of peak effect, a physician should not hesitate to repeat the

monitoring devices that incorporate capnography or pharyngeal air flow can detect apnea at the point of onset and should be used in hospitalized

common when opioids are administered with other CNS depressants. Co-administration of benzodiazepines is particularly likely to produce respiratory depression and should be avoided, especially in outpatient pain management. Because of this variability in patient response, initiation of therapy requires titration to optimal dose and interval. The most important principle is to provide adequate pain relief. This requires determining whether the drug has adequately relieved the pain and timely reassessment to determine the optimal interval for dosing. *The most common error made by physicians in managing severe pain with opioids is to prescribe an inadequate dose. Because many patients are reluctant to complain, this practice leads to needless suffering.* In the absence of sedation at the expected time of peak effect, a physician should not hesitate to repeat the initial dose to achieve satisfactory pain relief.

A now standard approach to the problem of achieving adequate pain relief is the use of patient-controlled analgesia (PCA). PCA uses a microprocessor-controlled infusion device that can deliver a baseline continuous dose of an opioid drug as well as preprogrammed additional doses whenever the patient pushes a button. The patient can then titrate the dose to the optimal level. This approach is used most extensively for the management of postoperative pain, but there is no reason why it should not be used for any hospitalized patient with persistent severe pain. PCA is also used for short-term home care of patients with intractable pain, such as that caused by metastatic cancer.

It is important to understand that the PCA device delivers small, repeated doses to maintain pain relief; in patients with severe pain, the pain must first be brought under control with a loading dose before transitioning to the PCA device. The bolus dose of the drug (typically 1 mg of morphine, 0.2 mg of hydromorphone, or 10 µg of fentanyl) can then be delivered repeatedly as needed. To prevent overdosing, PCA devices are programmed with a lockout period after each demand dose is delivered (typically starting at 10 min) and a limit on the total dose delivered per hour. Although some have advocated the use of a simultaneous continuous or basal infusion of the PCA drug, this may increase the risk of respiratory depression and has not been shown to increase the overall efficacy of the technique.

The availability of new routes of administration has extended the usefulness of opioid analgesics. Most important is the availability of spinal administration. Opioids can be infused through a spinal catheter placed either intrathecally or epidurally. By applying opioids directly to the spinal or epidural space adjacent to the spinal cord, regional analgesia can be obtained using relatively low total doses. Indeed, the dose required to produce effective analgesia when using morphine intrathecally (0.1–0.3 mg) is a fraction of that required to produce similar analgesia when administered intravenously (5–10 mg). In this way, side effects such as sedation, nausea, and respiratory depression can be minimized. This approach has been used extensively during labor and delivery and for postoperative pain relief following surgical procedures. Continuous intrathecal delivery via implanted spinal drug-delivery systems is now commonly used, particularly for the treatment of cancer-related pain that would require sedating doses for adequate pain control if given systemically. Opioids can also be given intranasally (butorphanol), rectally, and transdermally (fentanyl and buprenorphine), or through the oral mucosa (fentanyl), thus avoiding the discomfort of frequent injections in patients who cannot be given oral medication. The fentanyl and buprenorphine transdermal patches have the advantage of providing fairly steady plasma levels, which may improve patient comfort.

Recent additions to the armamentarium for treating opioid-induced side effects are the peripherally acting opioid antagonists alvimopan (Entereg) and methylnaltrexone (Rellistor). Alvimopan is available as an orally administered agent that is restricted to the intestinal lumen by limited absorption; methylnaltrexone is available in a subcutaneously administered form that has virtually no penetration into the CNS. Both agents act by binding to peripheral µ-receptors, thereby inhibiting or reversing the effects of opioids at these peripheral sites. The action of both agents is restricted to receptor sites outside of the CNS; thus, these drugs can reverse the adverse effects of opioid analgesics that are mediated through their peripheral receptors without reversing their CNS-mediated analgesic effects. Alvimopan has proven effective in lowering the duration of persistent ileus following abdominal surgery in patients receiving opioid analgesics for postoperative pain control. Methylnaltrexone has proven effective for relief of opioid-induced constipation in patients taking opioid analgesics on a chronic basis.

Opioid and COX Inhibitor Combinations

When used in combination, opioids and COX inhibitors have additive effects. Because a lower dose of each can be used to achieve the same degree of pain relief and their side effects are nonadditive, such combinations are used to lower the severity of dose-related side effects. However, fixed-ratio combinations of an opioid with acetaminophen carry an important risk. Dose escalation as a result of increased severity of pain or decreased opioid effect as a result of tolerance may lead to ingestion of levels of acetaminophen that are toxic to the liver. Although acetaminophen-related hepatotoxicity is uncommon, it remains a significant cause for liver failure. Thus, many practitioners have moved away from the use of opioid-acetaminophen combination analgesics to avoid the risk of excessive acetaminophen exposure as the dose of the analgesic is escalated.

TABLE 13-1

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

GENERIC NAME	DOSE, mg	INTERVAL	COMMENTS
Nonnarcotic Analgesics: Usual Doses and Intervals			

hepatotoxicity is uncommon, it remains a significant cause for liver failure. Thus, many practitioners have moved away from the use of opioid-

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 13-1

Drugs for Relief of Pain

GENERIC NAME	DOSE, mg	INTERVAL	COMMENTS
Nonnarcotic Analgesics: Usual Doses and Intervals			
Acetylsalicylic acid	650 PO	q4h	Enteric-coated preparations available
Acetaminophen	650 PO	q4h	Side effects uncommon
Ibuprofen	400 PO	q4–6h	Available without prescription
Naproxen	250–500 PO	q12h	Naproxen is the common NSAID that poses the least cardiovascular risk, but it has a somewhat higher incidence of gastrointestinal bleeding
Fenoprofen	200 PO	q4–6h	Contraindicated in renal disease
Indomethacin	25–50 PO	q8h	Gastrointestinal side effects common
Ketorolac	15–60 IM/IV	q4–6h	Available for parenteral use
Celecoxib	100–200 PO	q12–24h	Useful for arthritis
Valdecoxib	10–20 PO	q12–24h	Removed from U.S. market in 2005
GENERIC NAME	**PARENTERAL DOSE, mg**	**PO DOSE, mg**	**COMMENTS**
Narcotic Analgesics: Usual Doses and Intervals			
Codeine	30–60 q4h	30–60 q4h	Nausea common
Oxycodone	—	5–10 q4–6h	Usually available with acetaminophen or aspirin
Oxycodone extended-release	—	10-40 q12h	Oral extended-release tablet; high potential for misuse
Morphine	5 q4h	30 q4h	
Morphine sustained release	—	15–60 bid to tid	Oral slow-release preparation
Hydromorphone	1–2 q4h	2–4 q4h	Shorter acting than morphine sulfate
Levorphanol	2 q6–8h	4 q6–8h	Longer acting than morphine sulfate; absorbed well PO
Methadone	5–10 q6–8h	5–20 q6–8h	Due to long half-life, respiratory depression and sedation may persist after analgesic effect subsides; therapy should not be initiated with >40 mg/d, and dose escalation should be made no more frequently than every 3 days
Meperidine	50–100 q3–4h	300 q4h	Poorly absorbed PO; normeperidine is a toxic metabolite; routine use of this agent is not recommended
Fentanyl	25–100 µg/h	—	72-h transdermal patch

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Methadone	5–10 q6–8h	5–20 q6–8h	Due to long half-life, respiratory depression and sedation may persist after
Meperidine	50–100 q3–4h	300 q4h	Poorly absorbed PO; normeperidine is a toxic metabolite; routine use of this agent is not recommended
Butorphanol	—	1–2 q4h	Intranasal spray
Fentanyl	25–100 µg/h	—	72-h transdermal patch
Buprenorphine	5–20 µg/h		7-day transdermal patch
Buprenorphine	0.3 q6–8h		Parenteral administration
Tramadol	—	50–100 q4–6h	Mixed opioid/adrenergic action

GENERIC NAME	UPTAKE BLOCKADE		SEDATIVE POTENCY	ANTICHOLINERGIC POTENCY	ORTHOSTATIC HYPOTENSION	CARDIAC ARRHYTHMIA	AVERAGE DOSE, mg/d	RANGE, mg/d
	5-HT	NE						
Antidepressants[a]								
Doxepin	++	+	High	Moderate	Moderate	Less	200	75–400
Amitriptyline	++++	++	High	Highest	Moderate	Yes	150	25–300
Imipramine	++++	++	Moderate	Moderate	High	Yes	200	75–400
Nortriptyline	+++	++	Moderate	Moderate	Low	Yes	100	40–150
Desipramine	+++	++++	Low	Low	Low	Yes	150	50–300
Venlafaxine	+++	++	Low	None	None	No	150	75–400
Duloxetine	+++	+++	Low	None	None	No	40	30–60

GENERIC NAME	PO DOSE, mg	INTERVAL	COMMENTS
Anticonvulsants and Antiarrythmics[a]			
Carbamazepine	200–300	q6h	Rare aplastic anemia, GI irritation, hepatoitoxicity
Oxcarbamazepine	300	bid	Similar to carbamazepine
Gabapentin[b]	600–1200	q8h	Dizziness, GI irritation; useful in trigeminal neuralgia
Pregabalin	150–600	bid	Similar to gabapentin; dry mouth, edema

[a]Antidepressants, anticonvulsants, and antiarrhythmics have not been approved by the U.S. Food and Drug Administration (FDA) for the treatment of pain.

[b]Gabapentin in doses up to 1800 mg/d is FDA approved for postherpetic neuralgia.

Abbreviations: 5-HT, serotonin; NE, norepinephrine; NSAID, nonsteroidal anti-inflammatory agent.

CHRONIC PAIN

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

precipitating cause, e.g., fibromyalgia, or chronic headache. In many patients, chronic pain becomes a distinct disease unto itself. The pain generating mechanism is often difficult or impossible to determine with certainty; such patients are demanding of the physician's time and often appear

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

CHRONIC PAIN

Managing patients with chronic pain is intellectually and emotionally challenging. Sensitization of the nervous system can occur without an obvious precipitating cause, e.g., fibromyalgia, or chronic headache. In many patients, chronic pain becomes a distinct disease unto itself. The pain-generating mechanism is often difficult or impossible to determine with certainty; such patients are demanding of the physician's time and often appear emotionally distraught. The traditional medical approach of seeking an obscure organic pathology is often unhelpful. On the other hand, psychological evaluation and behaviorally based treatment paradigms are frequently helpful, particularly in the setting of a multidisciplinary pain-management center. Unfortunately, this approach, while effective, remains largely underused in current medical practice.

There are several factors that can cause, perpetuate, or exacerbate chronic pain. First, of course, the patient may simply have a disease that is characteristically painful for which there is presently no cure. Arthritis, cancer, chronic daily headaches, fibromyalgia, and diabetic neuropathy are examples of this. Second, there may be secondary perpetuating factors that are initiated by disease and persist after that disease has resolved. Examples include damaged sensory nerves, sympathetic efferent activity, and painful reflex muscle contraction (spasm). Finally, a variety of psychological conditions can exacerbate or even cause pain.

There are certain areas to which special attention should be paid in a patient's medical history. Because depression is the most common emotional disturbance in patients with chronic pain, patients should be questioned about their mood, appetite, sleep patterns, and daily activity. A simple standardized questionnaire, such as the Beck Depression Inventory, can be a useful screening device. It is important to remember that major depression is a common, treatable, and potentially fatal illness.

Other clues that a significant emotional disturbance is contributing to a patient's chronic pain complaint include pain that occurs in multiple, unrelated sites; a pattern of recurrent, but separate, pain problems beginning in childhood or adolescence; pain beginning at a time of emotional trauma, such as the loss of a parent or spouse; a history of physical or sexual abuse; and past or present substance abuse.

On examination, special attention should be paid to whether the patient guards the painful area and whether certain movements or postures are avoided because of pain. Discovering a mechanical component to the pain can be useful both diagnostically and therapeutically. Painful areas should be examined for deep tenderness, noting whether this is localized to muscle, ligamentous structures, or joints. Chronic myofascial pain is very common, and in these patients, deep palpation may reveal highly localized trigger points that are firm bands or knots in muscle. Relief of the pain following injection of local anesthetic into these trigger points supports the diagnosis. A neuropathic component to the pain is indicated by evidence of nerve damage, such as sensory impairment, exquisitely sensitive skin (allodynia), weakness, and muscle atrophy, or loss of deep tendon reflexes. Evidence suggesting sympathetic nervous system involvement includes the presence of diffuse swelling, changes in skin color and temperature, and hypersensitive skin and joint tenderness compared with the normal side. Relief of the pain with a sympathetic block supports the diagnosis, but once the condition becomes chronic, the response to sympathetic blockade is of variable magnitude and duration; the role for repeated sympathetic blocks in the overall management of CRPS is unclear.

A guiding principle in evaluating patients with chronic pain is to assess both emotional and somatic causal and perpetuating factors before initiating therapy. Addressing these issues together, rather than waiting to address emotional issues after somatic causes of pain have been ruled out, improves compliance in part because it assures patients that a psychological evaluation does not mean that the physician is questioning the validity of their complaint. Even when a somatic cause for a patient's pain can be found, it is still wise to look for other factors. For example, a cancer patient with painful bony metastases may have additional pain due to nerve damage and may also be depressed. Optimal therapy requires that each of these factors be assessed and treated.

TREATMENT

Chronic Pain

Once the evaluation process has been completed and the likely causative and exacerbating factors identified, an explicit treatment plan should be developed. An important part of this process is to identify specific and realistic functional goals for therapy, such as getting a good night's sleep, being able to go shopping, or returning to work. A multidisciplinary approach that uses medications, counseling, physical therapy, nerve blocks, and even surgery may be required to improve the patient's quality of life. There are also some newer, minimally invasive procedures that can be helpful for some patients with intractable pain. These include image-guided interventions such as epidural injection of glucocorticoids for acute radicular pain and radiofrequency treatment of the facet joints for chronic facet-related back and neck pain. For patients with severe and persistent pain that

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

patients who have not responded to conventional pharmacologic approaches. Referral to a multidisciplinary pain clinic for a full evaluation should precede any invasive procedure. Such referrals are clearly not necessary for all chronic pain patients. For some, pharmacologic management alone

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

for some patients with intractable pain. These include image-guided interventions such as epidural injection of glucocorticoids for acute radicular pain and radiofrequency treatment of the facet joints for chronic facet-related back and neck pain. For patients with severe and persistent pain that is unresponsive to more conservative treatment, placement of electrodes on peripheral nerves or within the spinal canal on nerve roots or in the space overlying the dorsal columns of the spinal cord (spinal cord stimulation) or implantation of intrathecal drug-delivery systems has shown significant benefit. The criteria for predicting which patients will respond to these procedures continue to evolve. They are generally reserved for patients who have not responded to conventional pharmacologic approaches. Referral to a multidisciplinary pain clinic for a full evaluation should precede any invasive procedure. Such referrals are clearly not necessary for all chronic pain patients. For some, pharmacologic management alone can provide adequate relief.

ANTIDEPRESSANT MEDICATIONS

The tricyclic antidepressants (TCAs), particularly nortriptyline and desipramine (Table 13-1), are useful for the management of chronic pain. Although developed for the treatment of depression, the TCAs have a spectrum of dose-related biologic activities that include analgesia in a variety of chronic clinical conditions. Although the mechanism is unknown, the analgesic effect of TCAs has a more rapid onset and occurs at a lower dose than is typically required for the treatment of depression. Furthermore, patients with chronic pain who are not depressed obtain pain relief with antidepressants. There is evidence that TCAs potentiate opioid analgesia, so they may be useful adjuncts for the treatment of severe persistent pain such as occurs with malignant tumors. **Table 13-2** lists some of the painful conditions that respond to TCAs. TCAs are of particular value in the management of neuropathic pain such as occurs in diabetic neuropathy and postherpetic neuralgia, for which there are few other therapeutic options.

The TCAs that have been shown to relieve pain have significant side effects (Table 13-1; **Chap. 452**). Some of these side effects, such as orthostatic hypotension, drowsiness, cardiac conduction delay, memory impairment, constipation, and urinary retention, are particularly problematic in elderly patients, and several are additive to the side effects of opioid analgesics. The selective serotonin reuptake inhibitors such as fluoxetine (Prozac) have fewer and less serious side effects than TCAs, but they are much less effective for relieving pain. It is of interest that venlafaxine (Effexor) and duloxetine (Cymbalta), which are nontricyclic antidepressants that block both serotonin and norepinephrine reuptake, appear to retain most of the pain-relieving effect of TCAs with a side effect profile more like that of the selective serotonin reuptake inhibitors. These drugs may be particularly useful in patients who cannot tolerate the side effects of TCAs.

ANTICONVULSANTS AND ANTIARRHYTHMICS

These drugs are useful primarily for patients with neuropathic pain. Phenytoin (Dilantin) and carbamazepine (Tegretol) were first shown to relieve the pain of trigeminal neuralgia **(Chap. 441)**. This pain has a characteristic brief, shooting, electric shock-like quality. In fact, anticonvulsants seem to be particularly helpful for pains that have such a lancinating quality. Newer anticonvulsants, the calcium channel alpha-2-delta subunit ligands gabapentin (Neurontin) and pregabalin (Lyrica), are effective for a broad range of neuropathic pains. Furthermore, because of their favorable side effect profile, these newer anticonvulsants are often used as first-line agents.

CANNABINOIDS

These agents are widely used for their analgesic properties, although published evidence suggests that any effects are likely to be modest, with small increases in pain threshold reported and variable reductions in clinical pain intensity. Cannabis more consistently reduces the unpleasantness of the pain experience and, in cancer-related pain, can lessen the nausea and vomiting associated with chemotherapy use. *Marijuana and related compounds are discussed in **Chap. 455**.*

CHRONIC OPIOID MEDICATION

The long-term use of opioids is accepted for patients with pain due to malignant disease. Although opioid use for chronic pain of nonmalignant origin is controversial, it is clear that, for many patients, opioids are the only option that produces meaningful pain relief. This is understandable because opioids are the most potent and have the broadest range of efficacy of any analgesic medications. Although addiction is rare in patients who first use opioids for pain relief, some degree of tolerance and physical dependence is likely with long-term use. Furthermore, studies suggest that long-term opioid therapy may worsen pain in some individuals, termed *opioid-induced hyperalgesia*. Therefore, before embarking on opioid therapy, other options should be explored, and the limitations and risks of opioids should be explained to the patient. It is also important to point out that some opioid analgesic medications have mixed agonist-antagonist properties (e.g., butorphanol and buprenorphine). From a practical standpoint, this means that they may worsen pain by inducing an abstinence syndrome in patients who are actively being treated with other opioids and are physically dependent.

With long-term outpatient use of orally administered opioids, it may be desirable to use long-acting compounds such as levorphanol, methadone,

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

formulations are approved primarily for patients who are already taking other opioids and should not be used as first-line opioids for pain. Although long-acting opioid preparations may provide superior pain relief in patients with a continuous pattern of ongoing pain, others suffer from

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

With long-term outpatient use of orally administered opioids, it may be desirable to use long-acting compounds such as levorphanol, methadone, extended-release morphine or oxycodone, or transdermal fentanyl (Table 13-1). The pharmacokinetic profiles of these drug preparations enable the maintenance of sustained analgesic blood levels, potentially minimizing side effects such as sedation that are associated with high peak plasma levels, and reducing the likelihood of rebound pain associated with a rapid fall in plasma opioid concentration. Extended-release opioid formulations are approved primarily for patients who are already taking other opioids and should not be used as first-line opioids for pain. Although long-acting opioid preparations may provide superior pain relief in patients with a continuous pattern of ongoing pain, others suffer from intermittent severe episodic pain and experience superior pain control and fewer side effects with the periodic use of short-acting opioid analgesics. Constipation is a virtually universal side effect of opioid use and should be treated expectantly. As noted earlier in the discussion of acute pain treatment, a recent advance for patients is the development of peripherally acting opioid antagonists that can reverse the constipation associated with opioid use without interfering with analgesia.

Soon after the introduction of an extended-release oxycodone formulation (OxyContin) in the late 1990s, a dramatic rise in emergency department visits and deaths associated with oxycodone ingestion appeared. This appears to be due primarily to individuals using a prescription opioid nonmedically. Drug-induced deaths have rapidly risen and are now the second leading cause of death in Americans, just behind motor vehicle fatalities. In 2011, the Office of National Drug Control Policy established a multifaceted approach to address prescription drug abuse, including prescription drug monitoring programs (PDMPs) that allow practitioners to determine if patients are receiving prescriptions from multiple providers and use of law enforcement to eliminate improper prescribing practices. In 2016, the Centers for Disease Control and Prevention (CDC) released the *CDC Guideline for Prescribing Opioids for Chronic Pain*, with recommendations for primary care clinicians who are prescribing opioids for chronic noncancer pain. A modified approach to opioid prescribing was published in 2019 by the Health and Human Services Task Force on chronic pain best medical practices. These guidelines address (1) when to initiate or continue opioids for chronic pain; (2) opioid selection, dosage, duration, follow-up, and discontinuation; and (3) assessing risk and addressing harms of opioid use. The recent increase in scrutiny leaves many practitioners hesitant to prescribe opioid analgesics, other than for brief periods to control pain associated with illness or injury. For now, the choice to begin chronic opioid therapy for a given patient is left to the individual practitioner. Pragmatic guidelines for properly selecting and monitoring patients receiving chronic opioid therapy are shown in **Table 13-3;** a checklist for primary care clinicians prescribing opioids for noncancer pain is shown in **Table 13-4**.

TREATMENT OF NEUROPATHIC PAIN

It is important to individualize treatment for patients with neuropathic pain. Several general principles should guide therapy: the first is to move quickly to provide relief, and the second is to minimize drug side effects. For example, in patients with postherpetic neuralgia and significant cutaneous hypersensitivity, topical lidocaine (Lidoderm patches) can provide immediate relief without side effects. The anticonvulsants gabapentin or pregabalin (see above) or antidepressants (nortriptyline, desipramine, duloxetine, or venlafaxine) can be used as first-line drugs for patients with neuropathic pain. Systemically administered antiarrhythmic drugs such as lidocaine and mexiletine are less likely to be effective. Although intravenous infusion of lidocaine can provide analgesia for patients with different types of neuropathic pain, the relief is usually transient, typically lasting just hours after the cessation of the infusion. The oral lidocaine congener mexiletine is poorly tolerated, producing frequent gastrointestinal adverse effects. There is no consensus on which class of drug should be used as a first-line treatment for any chronically painful condition. However, because relatively high doses of anticonvulsants are required for pain relief, sedation is not uncommon. Sedation is also a problem with TCAs but is much less of a problem with serotonin/norepinephrine reuptake inhibitors (SNRIs; e.g., venlafaxine and duloxetine). Thus, in the elderly or in patients whose daily activities require high-level mental activity, these drugs should be considered the first line. In contrast, opioid medications should be used as a second- or third-line drug class. Although highly effective for many painful conditions, opioids are sedating, and their effect tends to lessen over time, leading to dose escalation and, occasionally, a worsening of pain. A couple of interesting alternatives to pure opioids are two drugs with mixed opioid and norepinephrine reuptake action: tramadol and tapentadol. Tramadol is a relatively weak opioid but is sometimes effective for pain unresponsive to nonopioid analgesics. Tapentadol is a stronger opioid, but its analgesic action is apparently enhanced by the norepinephrine reuptake blockade. Similarly, drugs of different classes can be used in combination to optimize pain control. Repeated injection of botulinum toxin is an emerging approach that is showing some promise in treating focal neuropathic pain, particularly post-herpetic, trigeminal, and post-traumatic neuralgias.

It is worth emphasizing that many patients, especially those with chronic pain, seek medical attention primarily because they are suffering and because only physicians can provide the medications required for pain relief. A primary responsibility of all physicians is to minimize the physical and emotional discomfort of their patients. Familiarity with pain mechanisms and analgesic medications is an important step toward accomplishing this aim.

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Postherpetic neuralgiaa

TABLE 13-2

Painful Conditions That Respond to Tricyclic Antidepressants

Postherpetic neuralgia[a]
Diabetic neuropathy[a]
Fibromyalgia[a]
Tension headache[a]
Migraine headache[a]
Rheumatoid arthritis[a,b]
Chronic low back pain[b]
Cancer
Central poststroke pain

[a]Controlled trials demonstrate analgesia.

[b]Controlled studies indicate benefit but not analgesia.

TABLE 13-3

Guidelines for Selecting and Monitoring Patients Receiving Chronic Opioid Therapy (COT) for the Treatment of Chronic, Noncancer Pain

Patient Selection
• Conduct a history, physical examination, and appropriate testing, including an assessment of risk of substance abuse, misuse, or addiction. • Consider a trial of COT if pain is moderate or severe, pain is having an adverse impact on function or quality of life, and potential therapeutic benefits outweigh potential harms. • A benefit-to-harm evaluation, including a history, physical examination, and appropriate diagnostic testing, should be performed and documented before and on an ongoing basis during COT.
Informed Consent and Use of Management Plans
• Informed consent should be obtained. A continuing discussion with the patient regarding COT should include goals, expectations, potential risks, and alternatives to COT. • Consider using a written COT management plan to document patient and clinician responsibilities and expectations and assist in patient education.
Initiation and Titration
• Initial treatment with opioids should be considered as a therapeutic trial to determine whether COT is appropriate. • Opioid selection, initial dosing, and titration should be individualized according to the patient's health status, previous exposure to opioids, attainment of therapeutic goals, and predicted or observed harms.
• Reassess patients on COT periodically and as warranted by changing circumstances. Monitoring should include documentation of pain intensity and

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

controlled studies indicate benefit but not analgesia.

TABLE 13-3

Guidelines for Selecting and Monitoring Patients Receiving Chronic Opioid Therapy (COT) for the Treatment of Chronic, Noncancer Pain

Patient Selection
• Conduct a history, physical examination, and appropriate testing, including an assessment of risk of substance abuse, misuse, or addiction. • Consider a trial of COT if pain is moderate or severe, pain is having an adverse impact on function or quality of life, and potential therapeutic benefits outweigh potential harms. • A benefit-to-harm evaluation, including a history, physical examination, and appropriate diagnostic testing, should be performed and documented before and on an ongoing basis during COT.
Informed Consent and Use of Management Plans
• Informed consent should be obtained. A continuing discussion with the patient regarding COT should include goals, expectations, potential risks, and alternatives to COT. • Consider using a written COT management plan to document patient and clinician responsibilities and expectations and assist in patient education.
Initiation and Titration
• Initial treatment with opioids should be considered as a therapeutic trial to determine whether COT is appropriate. • Opioid selection, initial dosing, and titration should be individualized according to the patient's health status, previous exposure to opioids, attainment of therapeutic goals, and predicted or observed harms.
Monitoring
• Reassess patients on COT periodically and as warranted by changing circumstances. Monitoring should include documentation of pain intensity and level of functioning, assessments of progress toward achieving therapeutic goals, presence of adverse events, and adherence to prescribed therapies. • In patients on COT who are at high risk or who have engaged in aberrant drug-related behaviors, clinicians should periodically obtain urine drug screens or other information to confirm adherence to the COT plan of care. • In patients on COT not at high risk and not known to have engaged in aberrant drug-related behaviors, clinicians should consider periodically obtaining urine drug screens or other information to confirm adherence to the COT plan of care.

Source: Adapted with permission from R Chou, GJ Fanciullo, PG Fine et al: Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. J Pain 10:113-130, 2009. doi:10.1016/j.jpain.2008.10.008.

TABLE 13-4

Centers for Disease Control and Prevention Checklist for Prescribing Opioids for Chronic Pain

For Primary Care Providers Treating Adults (18+) with Chronic Pain ≥3 months, Excluding Cancer, Palliative, and End-of-Life Care
CHECKLIST
WHEN CONSIDERING LONG-TERM OPIOID THERAPY
• Set realistic goals for pain and function based on diagnosis (e.g., walk around the block). • Check that nonopioid therapies tried and optimized. • Discuss benefits and risks (e.g., addiction, overdose) with patient.   ○ Check urine drug screen. • Set criteria for stopping or continuing opioids.

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

10:113-130, 2009. doi:10.1016/j.jpain.2008.10.008.

TABLE 13-4

Centers for Disease Control and Prevention Checklist for Prescribing Opioids for Chronic Pain

For Primary Care Providers Treating Adults (18+) with Chronic Pain ≥3 months, Excluding Cancer, Palliative, and End-of-Life Care
CHECKLIST

WHEN CONSIDERING LONG-TERM OPIOID THERAPY
• Set realistic goals for pain and function based on diagnosis (e.g., walk around the block).
• Check that nonopioid therapies tried and optimized.
• Discuss benefits and risks (e.g., addiction, overdose) with patient.
• Evaluate risk of harm or misuse.
  ○ Discuss risk factors with patient.
  ○ Check prescription drug monitoring program (PDMP) data.
  ○ Check urine drug screen.
• Set criteria for stopping or continuing opioids.
• Assess baseline pain and function (e.g., Pain, Enjoyment, General Activity [PEG] scale).
• Schedule initial reassessment within 1–4 weeks.
• Prescribe short-acting opioids using lowest dosage on product labeling; match duration to scheduled reassessment.

IF RENEWING WITHOUT A PATIENT VISIT • Check that return visit is scheduled ≤3 months from last visit.

WHEN REASSESSING AT A PATIENT VISIT
• Continue opioids only after confirming clinically meaningful improvements in pain and function without significant risks or harm.
• Assess pain and function (e.g., PEG); compare results to baseline.
• Evaluate risk of harm or misuse:
  ○ Observe patient for signs of oversedation or overdose risk. If yes: Taper dose.
  ○ Check PDMP.
  ○ Check for opioid use disorder if indicated (e.g., difficulty controlling use). If yes: Refer for treatment.
• Check that nonopioid therapies optimized. Determine whether to continue, adjust, taper, or stop opioids.
• Calculate opioid dosage morphine milligram equivalent (MME).
  ○ If ≥50 MME/day total (≥50 mg hydrocodone; ≥33 mg oxycodone), increase frequency of follow-up; consider offering naloxone.
  ○ Avoid ≥90 MME/day total (≥90 mg hydrocodone; ≥60 mg oxycodone), or carefully justify; consider specialist referral.
• Schedule reassessment at regular intervals (≤3 months).

Source: Centers for Disease Control and Prevention, available at: https://stacks.cdc.gov/view/cdc/38025. Accessed May 25, 2017 (Public Domain).

FURTHER READING

De Vita MJ et al: Association of cannabinoid administration with experimental pain in healthy adults a systematic review and meta-analysis. JAMA Psychiatry 75:1118, 2018. [PubMed: 30422266]

Dowell D et al: CDC guideline for prescribing opioids for chronic pain—United States, 2016. JAMA 315:1624, 2016. [PubMed: 26977696]

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Sun EC et al: Incidence of and risk factors for chronic opioid use among opioid-naive patients in the postoperative period. JAMA Intern Med 176:1286, 2016. [PubMed: 27400458]

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

Finnerup NB et al: Pharmacotherapy for neuropathic pain in adults: A systematic review and meta-analysis. Lancet Neurol 14:162, 2015. [PubMed: 25575710]

Sun EC et al: Incidence of and risk factors for chronic opioid use among opioid-naive patients in the postoperative period. JAMA Intern Med 176:1286, 2016. [PubMed: 27400458]

U.S. Department of Health and Human Services: Pain management best practices inter-agency task force report: Updates, gaps, inconsistencies, and recommendations. May 2019.

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 13: Pain: Pathophysiology and Management, James P. Rathmell; Howard L. Fields
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 14: Chest Discomfort

David A. Morrow

INTRODUCTION

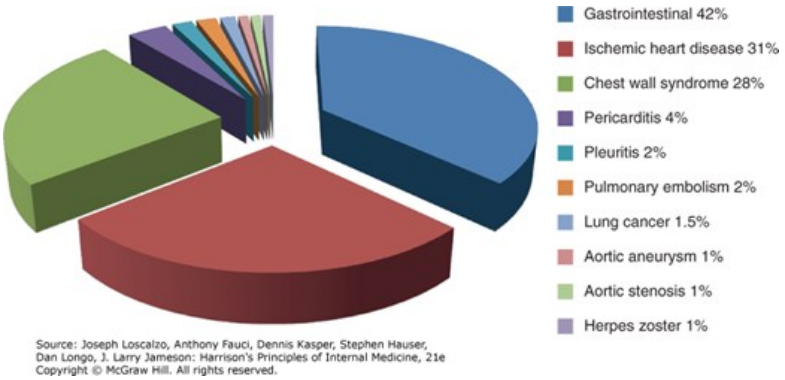
Chest discomfort is among the most common reasons for which patients present for medical attention at either an emergency department (ED) or an outpatient clinic. The evaluation of nontraumatic chest discomfort is inherently challenging owing to the broad variety of possible causes, a minority of which are life-threatening conditions that should not be missed. It is helpful to frame the initial diagnostic assessment and triage of patients with acute chest discomfort around three categories: (1) myocardial ischemia; (2) other cardiopulmonary causes (myopericardial disease, aortic emergencies, and pulmonary conditions); and (3) noncardiopulmonary causes. Although rapid identification of high-risk conditions is a priority of the initial assessment, strategies that incorporate routine liberal use of testing carry the potential for adverse effects of unnecessary investigations.

EPIDEMIOLOGY AND NATURAL HISTORY

Chest discomfort is one of the three most common reason for visits to the ED in the United States, resulting in 6 to 7 million emergency visits each year. More than 60% of patients with this presentation are hospitalized for further testing, and most of the remainder undergo additional investigation in the ED. Fewer than 15% of evaluated patients are eventually diagnosed with acute coronary syndrome (ACS), with rates of 10–20% in most series of unselected populations, and a rate as low as 5% in some studies. The most common diagnoses are gastrointestinal causes **(Fig. 14-1)**, and as few as 5% are other life-threatening cardiopulmonary conditions. In a large proportion of patients with transient acute chest discomfort, ACS or another acute cardiopulmonary cause is excluded but the cause is not determined. Therefore, the resources and time devoted to the evaluation of chest discomfort *in the absence of a severe cause* are substantial. Nevertheless, historically, a disconcerting 2–6% of patients with chest discomfort of presumed nonischemic etiology who are discharged from the ED were later deemed to have had a missed myocardial infarction (MI). Patients with a missed diagnosis of MI have a 30-day risk of death that is double that of their counterparts who are hospitalized.

FIGURE 14-1

Distribution of final discharge diagnoses in patients with nontraumatic acute chest pain. *(Figure prepared from data in P Fruergaard et al: Eur Heart J 17:1028, 1996.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The natural histories of ACS, myocarditis, acute pericardial diseases, pulmonary embolism, and aortic emergencies are discussed in **Chaps. 270**, **273**, **274**, **275**, **279**, and **280**, respectively. In a study of more than 350,000 patients with unspecified presumed noncardiopulmonary chest discomfort, the mortality rate 1 year after discharge was <2% and did not differ significantly from age-adjusted mortality in the general population. The estimated rate of major cardiovascular events through 30 days in patients with acute chest pain who had been stratified as low risk was 2.5% in a large population-based study that excluded patients with ST-segment elevation or definite noncardiac chest pain.

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The major etiologies of chest discomfort are discussed in this section and summarized in **Table 14-1**. Additional elements of the history, physical examination, and diagnostic testing that aid in distinguishing these causes are discussed in a later section (see "Approach to the Patient").

based study that excluded patients with ST-segment elevation or definite noncardiac chest pain.

CAUSES OF CHEST DISCOMFORT

The major etiologies of chest discomfort are discussed in this section and summarized in **Table 14-1**. Additional elements of the history, physical examination, and diagnostic testing that aid in distinguishing these causes are discussed in a later section (see "Approach to the Patient").

TABLE 14-1

Typical Clinical Features of Major Causes of Acute Chest Discomfort

SYSTEM	CONDITION	ONSET/DURATION	QUALITY	LOCATION	ASSOCIATED FEATURES
Cardiopulmonary					
Cardiac	Myocardial ischemia	*Stable angina:* Precipitated by exertion, cold, or stress; 2–10 min *Unstable angina:* Increasing pattern or at rest *Myocardial infarction:* Usually >30 min	Pressure, tightness, squeezing, heaviness, burning	Retrosternal; often radiation to neck, jaw, shoulders, or arms; sometimes epigastric	S_4 gallop or mitral regurgitation murmur (rare) during pain; S_3 or rales if severe ischemia or complication of myocardial infarction
	Pericarditis	Variable; hours to days; may be episodic	Pleuritic, sharp	Retrosternal or toward cardiac apex; may radiate to left shoulder	May be relieved by sitting up and leaning forward; pericardial friction rub
Vascular	Acute aortic syndrome	Sudden onset of unrelenting pain	Tearing or ripping; knifelike	Anterior chest, often radiating to back, between shoulder blades	Associated with hypertension and/or underlying connective tissue disorder; murmur of aortic insufficiency; loss of peripheral pulses
	Pulmonary embolism	Sudden onset	Pleuritic; may manifest as heaviness with massive pulmonary embolism	Often lateral, on the side of the embolism	Dyspnea, tachypnea, tachycardia, and hypotension
	Pulmonary hypertension	Variable; often exertional	Pressure	Substernal	Dyspnea, signs of increased venous pressure
Pulmonary	Pneumonia or pleuritis	Variable	Pleuritic	Unilateral, often localized	Dyspnea, cough, fever, rales, occasional rub
	Spontaneous pneumothorax	Sudden onset	Pleuritic	Lateral to side of pneumothorax	Dyspnea, decreased breath sounds on side of pneumothorax
Noncardiopulmonary					
Gastrointestinal	Esophageal reflux	10–60 min	Burning	Substernal, epigastric	Worsened by postprandial recumbency; relieved by antacids

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

	pneumothorax			pneumothorax	side of pneumothorax
Gastrointestinal	Esophageal reflux	10–60 min	Burning	Substernal, epigastric	Worsened by postprandial recumbency; relieved by antacids
	Esophageal spasm	2–30 min	Pressure, tightness, burning	Retrosternal	Can closely mimic angina
	Peptic ulcer	Prolonged; 60–90 min after meals	Burning	Epigastric, substernal	Relieved with food or antacids
	Gallbladder disease	Prolonged	Aching or colicky	Epigastric, right upper quadrant; sometimes to the back	May follow meal
Neuromuscular	Costochondritis	Variable	Aching	Sternal	Sometimes swollen, tender, warm over joint; may be reproduced by localized pressure on examination
	Cervical disk disease	Variable; may be sudden	Aching; may include numbness	Arms and shoulders	May be exacerbated by movement of neck
	Trauma or strain	Usually constant	Aching	Localized to area of strain	Reproduced by movement or palpation
	Herpes zoster	Usually prolonged	Sharp or burning	Dermatomal distribution	Vesicular rash in area of discomfort
Psychological	Emotional and psychiatric conditions	Variable; may be fleeting or prolonged	Variable; often manifests as tightness and dyspnea with feeling of panic or doom	Variable; may be retrosternal	Situational factors may precipitate symptoms; history of panic attacks, depression

MYOCARDIAL ISCHEMIA/INJURY

Myocardial ischemia causing chest discomfort, termed *angina pectoris*, is a primary clinical concern in patients presenting with chest symptoms. Myocardial ischemia is precipitated by an imbalance between myocardial oxygen requirements and myocardial oxygen supply, resulting in insufficient delivery of oxygen to meet the heart's metabolic demands. Myocardial oxygen consumption may be elevated by increases in heart rate, ventricular wall stress, and myocardial contractility, whereas myocardial oxygen supply is determined by coronary blood flow and coronary arterial oxygen content. When myocardial ischemia is sufficiently severe and prolonged in duration (as little as 20 min), irreversible cellular injury occurs, resulting in MI.

Ischemic heart disease is most commonly caused by atheromatous plaque that obstructs one or more of the epicardial coronary arteries. Stable ischemic heart disease **(Chap. 273)** usually results from the gradual atherosclerotic narrowing of the coronary arteries. *Stable angina* is characterized by ischemic episodes that are typically precipitated by a superimposed increase in oxygen demand during physical exertion and relieved upon resting. Ischemic heart disease becomes unstable, manifest by ischemia at rest or with an escalating pattern, most commonly when rupture or erosion of one or more atherosclerotic lesions triggers coronary thrombosis. Unstable ischemic heart disease is further classified clinically by the presence or absence of detectable acute myocardial injury and the presence or absence of ST-segment elevation on the patient's electrocardiogram (ECG). When acute coronary atherothrombosis occurs, the intracoronary thrombus may be partially obstructive, generally leading to myocardial ischemia in the absence of ST-segment elevation. Unstable ischemic heart disease is classified as *unstable angina* when there is no detectable acute myocardial injury and as *non–ST elevation MI* (NSTEMI) when there is evidence of acute myocardial necrosis **(Chap. 274)**. When the coronary thrombus is acutely and completely occlusive, transmural myocardial ischemia usually ensues, with ST-segment elevation on the ECG and myocardial necrosis leading to a diagnosis of *ST elevation MI* (STEMI; see **Chap. 275**).

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

acute coronary syndrome, which encompasses unstable angina, NSTEMI, and STEMI, is in general reserved for ischemia precipitated by acute coronary atherothrombosis. In order to guide therapeutic strategies, a standardized system for classification of MI has been expanded to discriminate MI

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

completely occlusive, transmural myocardial ischemia usually ensues, with ST-segment elevation on the ECG and myocardial necrosis leading to a diagnosis of *ST elevation MI* (STEMI; see **Chap. 275**).

Clinicians should be aware that unstable ischemic symptoms may also occur predominantly because of increased myocardial oxygen demand (e.g., during intense psychological stress or fever) or because of decreased oxygen delivery due to anemia, hypoxia, or hypotension. However, the term *acute coronary syndrome*, which encompasses unstable angina, NSTEMI, and STEMI, is in general reserved for ischemia precipitated by acute coronary atherothrombosis. In order to guide therapeutic strategies, a standardized system for classification of MI has been expanded to discriminate MI resulting from acute coronary thrombosis (type 1 MI) from MI occurring secondary to other imbalances of myocardial oxygen supply and demand (type 2 MI; see **Chap. 274**). These conditions are additionally distinguished from nonischemic causes of acute myocardial injury, such as myocarditis.

Other contributors to stable and unstable ischemic heart disease, such as endothelial dysfunction, microvascular disease, and vasospasm, may exist alone or in combination with coronary atherosclerosis and may be the dominant cause of myocardial ischemia in some patients. Moreover, nonatherosclerotic processes, including congenital abnormalities of the coronary vessels, myocardial bridging, coronary arteritis, and radiation-induced coronary disease, can lead to coronary obstruction. In addition, conditions associated with extreme myocardial oxygen demand and impaired endocardial blood flow, such as aortic valve disease **(Chap. 280)**, hypertrophic cardiomyopathy, or idiopathic dilated cardiomyopathy **(Chap. 259)**, can precipitate myocardial ischemia in patients with or without underlying obstructive atherosclerosis.

Characteristics of Ischemic Chest Discomfort

The clinical characteristics of angina pectoris, often referred to simply as "angina," are highly similar whether the ischemic discomfort is a manifestation of stable ischemic heart disease, unstable angina, or MI; the exceptions are differences in the pattern and duration of symptoms associated with these syndromes (Table 14-1). Heberden initially described angina as a sense of "strangling and anxiety." Chest discomfort characteristic of myocardial ischemia is typically described as aching, heavy, squeezing, crushing, or constricting. However, in a substantial minority of patients, the quality of discomfort is extremely vague and may be described as a mild tightness, or merely an uncomfortable feeling, that sometimes is experienced as numbness or a burning sensation. The site of the discomfort is usually retrosternal, but radiation is common and generally occurs down the ulnar surface of the left arm; the right arm, both arms, neck, jaw, or shoulders may also be involved. These and other characteristics of ischemic chest discomfort pertinent to discrimination from other causes of chest pain are discussed later in this chapter (see "Approach to the Patient").

Stable angina usually begins gradually and reaches its maximal intensity over a period of minutes before dissipating within several minutes with rest or with nitroglycerin. The discomfort typically occurs predictably at a characteristic level of exertion or psychological stress. By definition, unstable angina is manifest by anginal chest discomfort that occurs with progressively lower intensity of physical activity or even at rest. Chest discomfort associated with MI is commonly more severe, is prolonged (usually lasting ≥30 min), and is not relieved by rest.

Mechanisms of Cardiac Pain

The neural pathways involved in ischemic cardiac pain are poorly understood. Ischemic episodes are thought to excite local chemosensitive and mechanoreceptive receptors that, in turn, stimulate release of adenosine, bradykinin, and other substances that activate the sensory ends of sympathetic and vagal afferent fibers. The afferent fibers traverse the nerves that connect to the upper five thoracic sympathetic ganglia and upper five distal thoracic roots of the spinal cord. From there, impulses are transmitted to the thalamus. Within the spinal cord, cardiac sympathetic afferent impulses may converge with impulses from somatic thoracic structures, and this convergence may be the basis for referred cardiac pain. In addition, cardiac vagal afferent fibers synapse in the nucleus tractus solitarius of the medulla and then descend to the upper cervical spinothalamic tract, and this route may contribute to anginal pain experienced in the neck and jaw.

OTHER CARDIOPULMONARY CAUSES

Pericardial and Other Myocardial Diseases

Inflammation of the pericardium due to infectious or noninfectious causes can be responsible for acute or chronic chest discomfort **(See also Chap. 270)**. The visceral surface and most of the parietal surface of the pericardium are insensitive to pain. Therefore, the pain of pericarditis is thought to arise principally from associated pleural inflammation. Because of this pleural association, the discomfort of pericarditis is usually pleuritic pain that is exacerbated by breathing, coughing, or changes in position. Moreover, owing to the overlapping sensory supply of the central diaphragm via the phrenic nerve with somatic sensory fibers originating in the third to fifth cervical segments, the pain of pleural and pericardial inflammation is often

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

varied. Chest discomfort may either originate with inflammatory injury of the myocardium or be due to severe increases in wall stress related to poor ventricular performance. The symptoms of *Takotsubo (stress-related) cardiomyopathy* often start abruptly with chest pain and shortness of breath.

phrenic nerve with somatic sensory fibers originating in the third to fifth cervical segments, the pain of pleural and pericardial inflammation is often referred to the shoulder and neck. Involvement of the pleural surface of the lateral diaphragm can lead to pain in the upper abdomen.

Acute inflammatory and other nonischemic myocardial diseases can also produce chest discomfort. The symptoms of acute myocarditis are highly varied. Chest discomfort may either originate with inflammatory injury of the myocardium or be due to severe increases in wall stress related to poor ventricular performance. The symptoms of *Takotsubo (stress-related) cardiomyopathy* often start abruptly with chest pain and shortness of breath. This form of cardiomyopathy, in its most recognizable form, is triggered by an emotionally or physically stressful event and may mimic acute MI because of its commonly associated ECG abnormalities, including ST-segment elevation, and elevated biomarkers of myocardial injury. Observational studies support a predilection for women >50 years of age.

Diseases of the Aorta

Acute aortic dissection (Fig. 14-1) is a less common cause of chest discomfort but is important because of the catastrophic natural history of certain subsets of cases when recognized late or left untreated **(See also Chap. 280)**. Acute aortic syndromes encompass a spectrum of acute aortic diseases related to disruption of the media of the aortic wall. *Aortic dissection* involves a tear in the aortic intima, resulting in separation of the media and creation of a separate "false" lumen. A *penetrating ulcer* has been described as ulceration of an aortic atheromatous plaque that extends through the intima and into the aortic media, with the potential to initiate an intramedial dissection or rupture into the adventitia. *Intramural hematoma* is an aortic wall hematoma with no demonstrable intimal flap, no radiologically apparent intimal tear, and no false lumen. Intramural hematoma can occur due to either rupture of the vasa vasorum or, less commonly, a penetrating ulcer.

Each of these subtypes of acute aortic syndrome typically presents with chest discomfort that is often severe, sudden in onset, and sometimes described as "tearing" in quality. Acute aortic syndromes involving the *ascending* aorta tend to cause pain in the midline of the anterior chest, whereas *descending* aortic syndromes most often present with pain in the back. Therefore, dissections that begin in the ascending aorta and extend to the descending aorta tend to cause pain in the front of the chest that extends toward the back, between the shoulder blades. Proximal aortic dissections that involve the ascending aorta (type A in the Stanford nomenclature) are at high risk for major complications that may influence the clinical presentation, including (1) compromise of the aortic ostia of the coronary arteries, resulting in MI; (2) disruption of the aortic valve, causing acute aortic insufficiency; and (3) rupture of the hematoma into the pericardial space, leading to pericardial tamponade.

Knowledge of the epidemiology of acute aortic syndromes can be helpful in maintaining awareness of this relatively uncommon group of disorders (estimated annual incidence, 3 cases per 100,000 population). Nontraumatic aortic dissections are very rare in the absence of hypertension or conditions associated with deterioration of the elastic or muscular components of the aortic media, including pregnancy, bicuspid aortic disease, or inherited connective tissue diseases, such as Marfan and Ehlers-Danlos syndromes.

Although aortic aneurysms are most often asymptomatic, thoracic aortic aneurysms can cause chest pain and other symptoms by compressing adjacent structures. This pain tends to be steady, deep, and occasionally severe. Aortitis, whether of noninfectious or infectious etiology, in the absence of aortic dissection is a rare cause of chest or back discomfort.

Pulmonary Conditions

Pulmonary and pulmonary-vascular conditions that cause chest discomfort usually do so in conjunction with dyspnea and often produce symptoms that have a pleuritic nature.

PULMONARY EMBOLISM

Pulmonary emboli (annual incidence, ~1 per 1000) can produce dyspnea and chest discomfort that is sudden in onset (SEE ALSO CHAP. 279). Typically pleuritic in pattern, the chest discomfort associated with pulmonary embolism may result from (1) involvement of the pleural surface of the lung adjacent to a resultant pulmonary infarction; (2) distention of the pulmonary artery; or (3) possibly, right ventricular wall stress and/or subendocardial ischemia related to acute pulmonary hypertension. The pain associated with small pulmonary emboli is often lateral and pleuritic and is believed to be related to the first of these three possible mechanisms. In contrast, massive pulmonary emboli may cause severe substernal pain that may mimic an MI and that is plausibly attributed to the second and third of these potential mechanisms. Massive or submassive pulmonary embolism may also be associated with syncope, hypotension, and signs of right heart failure. Other typical characteristics that aid in the recognition of pulmonary embolism are discussed later in this chapter (see "Approach to the Patient").

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

symptoms are usually sudden in onset, and dyspnea may be mild; thus, presentation to medical attention is sometimes delayed. *Secondary*

Primary spontaneous pneumothorax is a rare cause of chest discomfort, with an estimated annual incidence in the United States of 7 per 100,000 among men and <2 per 100,000 among women **(SEE ALSO CHAP. 294)**. Risk factors include male sex, smoking, family history, and Marfan syndrome. The symptoms are usually sudden in onset, and dyspnea may be mild; thus, presentation to medical attention is sometimes delayed. *Secondary spontaneous pneumothorax* may occur in patients with underlying lung disorders, such as chronic obstructive pulmonary disease, asthma, or cystic fibrosis, and usually produces symptoms that are more severe. Tension pneumothorax is a medical emergency caused by trapped intrathoracic air that precipitates hemodynamic collapse.

Other Pulmonary Parenchymal, Pleural, or Vascular Disease

Most pulmonary diseases that produce chest pain, including pneumonia and malignancy, do so because of involvement of the pleura or surrounding structures **(See also Chaps. 283, 284, and 294)**. Pleurisy is typically described as a knifelike pain that is worsened by inspiration or coughing. In contrast, chronic pulmonary hypertension can manifest as chest pain that may be very similar to angina in its characteristics, suggesting right ventricular myocardial ischemia in some cases. Reactive airways diseases similarly can cause chest tightness associated with breathlessness rather than pleurisy.

NONCARDIOPULMONARY CAUSES

Gastrointestinal Conditions

Gastrointestinal disorders are the most common cause of nontraumatic chest discomfort and often produce symptoms that are difficult to discern from more serious causes of chest pain, including myocardial ischemia **(See also Chap. 321)**. Esophageal disorders, in particular, may simulate angina in the character and location of the pain. Gastroesophageal reflux and disorders of esophageal motility are common and should be considered in the differential diagnosis of chest pain (Fig. 14-1 and Table 14-1). The pain of esophageal spasm is commonly an intense, squeezing discomfort that is retrosternal in location and, like angina, may be relieved by nitroglycerin or dihydropyridine calcium channel antagonists. Chest pain can also result from injury to the esophagus, such as a Mallory-Weiss tear or even an esophageal rupture (Boerhaave's syndrome) caused by severe vomiting. Peptic ulcer disease is most commonly epigastric in location but can radiate into the chest (Table 14-1).

Hepatobiliary disorders, including cholecystitis and biliary colic, may mimic acute cardiopulmonary diseases. Although the pain arising from these disorders usually localizes to the right upper quadrant of the abdomen, it is variable and may be felt in the epigastrium and radiate to the back and lower chest. This discomfort is sometimes referred to the scapula or may in rare cases be felt in the shoulder, suggesting diaphragmatic irritation. The pain is steady, usually lasts several hours, and subsides spontaneously, without symptoms between attacks. Pain resulting from pancreatitis is typically aching epigastric pain that radiates to the back.

Musculoskeletal and Other Causes

Chest discomfort can be produced by any musculoskeletal disorder involving the chest wall or the nerves of the chest wall, neck, or upper limbs **(See also Chap. 360)**. Costochondritis causing tenderness of the costochondral junctions (*Tietze's syndrome*) is relatively common. Cervical radiculitis may manifest as a prolonged or constant aching discomfort in the upper chest and limbs. The pain may be exacerbated by motion of the neck. Occasionally, chest pain can be caused by compression of the brachial plexus by the cervical ribs, and tendinitis or bursitis involving the left shoulder may mimic the radiation of angina. Pain in a dermatomal distribution can also be caused by cramping of intercostal muscles or by herpes zoster **(Chap. 193)**.

Emotional and Psychiatric Conditions

As many as 10% of patients who present to EDs with acute chest discomfort have a panic disorder or related condition (Table 14-1). The symptoms may include chest tightness or aching that is associated with a sense of anxiety and difficulty breathing. The symptoms may be prolonged or fleeting.

APPROACH TO THE PATIENT

Chest Discomfort

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

cardiopulmonary processes, including ACS, acute aortic syndrome, pulmonary embolism, tension pneumothorax, and pericarditis with tamponade.

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Given the broad set of potential causes and the heterogeneous risk of serious complications in patients who present with acute nontraumatic chest discomfort, the priorities of the initial clinical encounter include assessment of (1) the patient's clinical stability and (2) the probability that the patient has an underlying cause of the discomfort that may be life-threatening. The high-risk conditions of principal concern are acute cardiopulmonary processes, including ACS, acute aortic syndrome, pulmonary embolism, tension pneumothorax, and pericarditis with tamponade. Fulminant myocarditis also carries a poor prognosis but is usually also manifest by heart failure symptoms. Among noncardiopulmonary causes of chest pain, esophageal rupture likely holds the greatest urgency for diagnosis. Patients with these conditions may deteriorate rapidly despite initially appearing well. The remaining population with noncardiopulmonary conditions has a more favorable prognosis during completion of the diagnostic workup. A rapid targeted assessment for a serious cardiopulmonary cause is of particular relevance for patients with acute ongoing pain who have presented for emergency evaluation. Among patients presenting in the outpatient setting with chronic pain or pain that has resolved, a general diagnostic assessment is reasonably undertaken (see "Outpatient Evaluation of Chest Discomfort," below). A series of questions that can be used to structure the clinical evaluation of patients with chest discomfort is shown in **Table 14-2**.

HISTORY

The evaluation of nontraumatic chest discomfort relies heavily on the clinical history and physical examination to direct subsequent diagnostic testing. The evaluating clinician should assess the quality, location (including radiation), and pattern (including onset and duration) of the pain as well as any provoking or alleviating factors. The presence of associated symptoms may also be useful in establishing a diagnosis.

Quality of Pain

The quality of chest discomfort alone is never sufficient to establish a diagnosis. However, the characteristics of the pain are pivotal in formulating an initial clinical impression and assessing the likelihood of a serious cardiopulmonary process (Table 14-1), including ACS in particular **(Fig. 14-2)**. Pressure or tightness is consistent with a typical presentation of myocardial ischemic pain. Nevertheless, the clinician must remember that some patients with ischemic chest symptoms deny any "pain" but rather complain of dyspnea or a vague sense of anxiety. The severity of the discomfort has poor diagnostic accuracy. It is often helpful to ask about the similarity of the discomfort to previous definite ischemic symptoms. It is unusual for angina to be sharp, as in knifelike, stabbing, or pleuritic; however, patients sometimes use the word "sharp" to convey the intensity of discomfort rather than the quality. Pleuritic discomfort is suggestive of a process involving the pleura, including pericarditis, pulmonary embolism, or pulmonary parenchymal processes. Less frequently, the pain of pericarditis or massive pulmonary embolism is a steady severe pressure or aching that can be difficult to discriminate from myocardial ischemia. "Tearing" or "ripping" pain is often described by patients with acute aortic dissection. However, acute aortic emergencies also present commonly with knifelike pain. A burning quality can suggest acid reflux or peptic ulcer disease but may also occur with myocardial ischemia. Esophageal pain, particularly with spasm, can be a severe squeezing discomfort identical to angina.

Location of Discomfort

A substernal location with radiation to the neck, jaw, shoulder, or arms is typical of myocardial ischemic discomfort. Radiation to both arms has a particularly high association with MI as the etiology. Some patients present with aching in sites of radiated pain as their only symptoms of ischemia. However, pain that is highly localized—e.g., that which can be demarcated by the tip of one finger—is highly unusual for angina. A retrosternal location should prompt consideration of esophageal pain; however, other gastrointestinal conditions usually present with pain that is most intense in the abdomen or epigastrium, with possible radiation into the chest. Angina may also occur in an epigastric location. Pain that occurs solely above the mandible or below the epigastrium is rarely angina. Severe pain radiating to the back, particularly between the shoulder blades, should prompt consideration of an acute aortic syndrome. Radiation to the trapezius ridge is characteristic of pericardial pain and does not usually occur with angina.

Pattern

Myocardial ischemic discomfort usually builds over minutes and is exacerbated by activity and mitigated by rest. In contrast, pain that reaches its peak intensity immediately is more suggestive of aortic dissection, pulmonary embolism, or spontaneous pneumothorax. Pain that is fleeting (lasting only a few seconds) is rarely ischemic in origin. Similarly, pain that is constant in intensity for a prolonged period (many hours to days) is unlikely to represent myocardial ischemia if it occurs in the absence of other clinical consequences, such as abnormalities of the ECG, elevation of cardiac biomarkers, or clinical sequelae (e.g., heart failure or hypotension). Both myocardial ischemia and acid reflux may have their onset in the morning.

Provoking and Alleviating Factors

Patients with myocardial ischemic pain usually prefer to rest, sit, or stop walking. However, clinicians should be aware of the phenomenon of

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

leaning forward. Gastroesophageal reflux may be exacerbated by alcohol, some foods, or a reclined position. Relief can occur with sitting.

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Patients with myocardial ischemic pain usually prefer to rest, sit, or stop walking. However, clinicians should be aware of the phenomenon of "warm-up angina" in which some patients experience relief of angina as they continue at the same or even a greater level of exertion **(Chap. 273)**. Alterations in the intensity of pain with changes in position or movement of the upper extremities and neck are less likely with myocardial ischemia and suggest a musculoskeletal etiology. The pain of pericarditis, however, often is worse in the supine position and relieved by sitting upright and leaning forward. Gastroesophageal reflux may be exacerbated by alcohol, some foods, or a reclined position. Relief can occur with sitting.

Exacerbation by eating suggests a gastrointestinal etiology such as peptic ulcer disease, cholecystitis, or pancreatitis. Peptic ulcer disease tends to become symptomatic 60–90 min after meals. However, in the setting of severe coronary atherosclerosis, redistribution of blood flow to the splanchnic vasculature after eating can trigger postprandial angina. The discomfort of acid reflux and peptic ulcer disease is usually diminished promptly by acid-reducing therapies. In contrast with its impact in some patients with angina, physical exertion is very unlikely to alter symptoms from gastrointestinal causes of chest pain. Relief of chest discomfort within minutes after administration of nitroglycerin is suggestive of but not sufficiently sensitive or specific for a definitive diagnosis of myocardial ischemia. Esophageal spasm may also be relieved promptly with nitroglycerin. A delay of >10 min before relief is obtained after nitroglycerin suggests that the symptoms either are not caused by ischemia or are caused by severe ischemia, such as during acute MI.

Associated Symptoms

Symptoms that accompany myocardial ischemia may include diaphoresis, dyspnea, nausea, fatigue, faintness, and eructations. In addition, these symptoms may exist in isolation as anginal equivalents (i.e., symptoms of myocardial ischemia other than typical angina), particularly in women and the elderly. Dyspnea may occur with multiple conditions considered in the differential diagnosis of chest pain and thus is not discriminative, but the presence of dyspnea is important because it suggests a cardiopulmonary etiology. Sudden onset of significant respiratory distress should lead to consideration of pulmonary embolism and spontaneous pneumothorax. Hemoptysis may occur with pulmonary embolism or as blood-tinged frothy sputum in severe heart failure but usually points toward a pulmonary parenchymal etiology of chest symptoms. Presentation with syncope or presyncope should prompt consideration of hemodynamically significant pulmonary embolism or aortic dissection as well as ischemic arrhythmias. Although nausea and vomiting suggest a gastrointestinal disorder, these symptoms may occur in the setting of MI (more commonly inferior MI), presumably because of activation of the vagal reflex or stimulation of left ventricular receptors as part of the Bezold-Jarisch reflex.

Past Medical History

The past medical history is useful in assessing the patient for risk factors for coronary atherosclerosis and venous thromboembolism **(Chap. 279)** as well as for conditions that may predispose the patient to specific disorders. For example, a history of connective tissue diseases such as Marfan syndrome should heighten the clinician's suspicion of an acute aortic syndrome or spontaneous pneumothorax. A careful history may elicit clues about depression or prior panic attacks.

PHYSICAL EXAMINATION

In addition to providing an initial assessment of the patient's clinical stability, the physical examination of patients with chest discomfort can provide direct evidence of specific etiologies of chest pain (e.g., unilateral absence of lung sounds) and can identify potential precipitants of acute cardiopulmonary causes of chest pain (e.g., uncontrolled hypertension), relevant comorbid conditions (e.g., obstructive pulmonary disease), and complications of the presenting syndrome (e.g., heart failure). However, because the findings on physical examination may be normal in patients with unstable ischemic heart disease, an unremarkable physical exam is not definitively reassuring.

General

The patient's general appearance is helpful in establishing an initial impression of the severity of illness. Patients with acute MI or other acute cardiopulmonary disorders often appear anxious, uncomfortable, pale, cyanotic, or diaphoretic. Patients who are massaging or clutching their chests may describe their pain with a clenched fist held against the sternum (*Levine's sign*). Occasionally, body habitus is helpful—e.g., in patients with Marfan syndrome or the prototypical young, tall, thin man with spontaneous pneumothorax.

Vital Signs

Significant tachycardia and hypotension are indicative of important hemodynamic consequences of the underlying cause of chest discomfort and should prompt a rapid survey for the most severe conditions, such as acute MI with cardiogenic shock, massive pulmonary embolism, pericarditis with tamponade, or tension pneumothorax. Acute aortic emergencies usually present with severe hypertension but may be associated with profound hypotension when there is coronary arterial compromise or dissection into the pericardium. Sinus tachycardia is an important manifestation of submassive pulmonary embolism. Tachypnea and hypoxemia point toward a pulmonary cause. The presence of low-grade fever is nonspecific because it may occur with MI and with thromboembolism in addition to infection.

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ventricular dysfunction from severe ischemia/infarction as well as acute valvular complications of MI or aortic dissection can lead to pulmonary edema, which is an indicator of high risk.

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

nonspecific because it may occur with MI and with thromboembolism in addition to infection.

Pulmonary

Examination of the lungs may localize a primary pulmonary cause of chest discomfort, as in cases of pneumonia, asthma, or pneumothorax. Left ventricular dysfunction from severe ischemia/infarction as well as acute valvular complications of MI or aortic dissection can lead to pulmonary edema, which is an indicator of high risk.

Cardiac

The jugular venous pulse is often normal in patients with acute myocardial ischemia but may reveal characteristic patterns with pericardial tamponade or acute right ventricular dysfunction **(Chaps. 239 and 270)**. Cardiac auscultation may reveal a third or, more commonly, a fourth heart sound, reflecting myocardial systolic or diastolic dysfunction. Murmurs of mitral regurgitation or a ventricular-septal defect may indicate mechanical complications of STEMI. A murmur of aortic insufficiency may be a complication of ascending aortic dissection. Other murmurs may reveal underlying cardiac disorders contributory to ischemia (e.g., aortic stenosis or hypertrophic cardiomyopathy). Pericardial friction rubs reflect pericardial inflammation.

Abdominal

Localizing tenderness on the abdominal exam is useful in identifying a gastrointestinal cause of the presenting syndrome. Abdominal findings are infrequent with purely acute cardiopulmonary problems, except in the case of right-sided heart failure leading to hepatic congestion.

Extremities

Vascular pulse deficits may reflect underlying chronic atherosclerosis, which increases the likelihood of coronary artery disease. However, evidence of acute limb ischemia with loss of the pulse and pallor, particularly in the upper extremities, can indicate catastrophic consequences of aortic dissection. Unilateral lower-extremity swelling should raise suspicion about venous thromboembolism.

Musculoskeletal

Pain arising from the costochondral and chondrosternal articulations may be associated with localized swelling, redness, or marked localized tenderness. Pain on palpation of these joints is usually well localized and is a useful clinical sign, although deep palpation may elicit pain in the absence of costochondritis. Although palpation of the chest wall often elicits pain in patients with various musculoskeletal conditions, it should be appreciated that chest wall tenderness does not exclude myocardial ischemia. Sensory deficits in the upper extremities may be indicative of cervical disk disease.

ELECTROCARDIOGRAPHY

Electrocardiography is crucial in the evaluation of nontraumatic chest discomfort. The ECG is pivotal for identifying patients with ongoing ischemia as the principal reason for their presentation as well as secondary cardiac complications of other disorders. Professional society guidelines recommend that an ECG be obtained within 10 min of presentation, with the primary goal of identifying patients with ST-segment elevation diagnostic of MI who are candidates for immediate interventions to restore flow in the occluded coronary artery. ST-segment depression and symmetric T-wave inversions at least 0.2 mV in depth are useful for detecting myocardial ischemia in the absence of STEMI and are also indicative of higher risk of death or recurrent ischemia. Serial performance of ECGs (every 30–60 min) is recommended in the ED evaluation of suspected ACS. In addition, an ECG with right-sided lead placement should be considered in patients with clinically suspected ischemia and a nondiagnostic standard 12-lead ECG. Despite the value of the resting ECG, its sensitivity for ischemia is poor—as low as 20% in some studies.

Abnormalities of the ST segment and T wave may occur in a variety of conditions, including pulmonary embolism, ventricular hypertrophy, acute and chronic pericarditis, myocarditis, electrolyte imbalance, and metabolic disorders. Notably, hyperventilation associated with panic disorder can also lead to nonspecific ST and T-wave abnormalities. Pulmonary embolism is most often associated with sinus tachycardia but can also lead to rightward shift of the ECG axis, manifesting as an S-wave in lead I, with a Q-wave and T-wave in lead III **(Chaps. 240 and 279)**. In patients with ST-segment elevation, the presence of diffuse lead involvement not corresponding to a specific coronary anatomic distribution and PR-segment depression can aid in distinguishing pericarditis from acute MI.

CHEST RADIOGRAPHY

(See Chap. A12) Plain radiography of the chest is performed routinely when patients present with acute chest discomfort and selectively when individuals who are being evaluated as outpatients have subacute or chronic pain. The chest radiograph is most useful for identifying pulmonary processes, such as pneumonia or pneumothorax. Findings are often unremarkable in patients with ACS, but pulmonary edema may be evident. Other specific findings include widening of the mediastinum in some patients with aortic dissection, Hampton's hump or Westermark's sign in

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Laboratory testing in patients with acute chest pain is focused on the detection of myocardial injury. Such injury can be detected by the presence of circulating proteins released from damaged cardiomyocytes. Owing to the time necessary for this release, initial biomarkers of injury may be in the

processes, such as pneumonia or pneumothorax. Findings are often unremarkable in patients with ACS, but pulmonary edema may be evident. Other specific findings include widening of the mediastinum in some patients with aortic dissection, Hampton's hump or Westermark's sign in patients with pulmonary embolism (**Chaps. 279** and **A12**), or pericardial calcification in chronic pericarditis.

CARDIAC BIOMARKERS

Laboratory testing in patients with acute chest pain is focused on the detection of myocardial injury. Such injury can be detected by the presence of circulating proteins released from damaged cardiomyocytes. Owing to the time necessary for this release, initial biomarkers of injury may be in the normal range, even in patients with STEMI. Cardiac troponin is the preferred biomarker for the diagnosis of MI and should be measured in all patients with suspected ACS. It is not necessary or advisable to measure troponin in patients without suspicion of ACS unless this test is being used specifically for risk stratification (e.g., in pulmonary embolism or heart failure).

The development of cardiac troponin assays with progressively greater analytical sensitivity has facilitated detection of substantially lower blood concentrations of troponin than was previously possible. This evolution permits earlier detection of myocardial injury and more reliable discrimination of changing values, enhances the overall accuracy of a diagnosis of MI, and improves risk stratification in suspected ACS. For these reasons, high-sensitivity assays are generally preferred over prior generation troponin assays. The greater negative predictive value of a negative troponin result with high-sensitivity assays is an advantage in the evaluation of chest pain in the ED. Rapid rule-out protocols that use serial testing and changes in troponin concentration over as short a period as 1–2 h appear to perform well for diagnosis of ACS when using a high-sensitivity troponin assay. Troponin should be measured at presentation and repeated at 1–3 h using high-sensitivity troponin and 3–6 h using conventional troponin assays. Additional troponin measurements may be warranted beyond 3–6 h when the clinical condition still suggests possible ACS or if there is diagnostic uncertainty. In patients presenting more than 2–3 h after symptom onset, a concentration of cardiac troponin, at the time of hospital presentation, below the limit of detection using a high-sensitivity assay may be sufficient to exclude MI with a negative predictive value >99%.

With the use of high-sensitivity assays for troponin, myocardial injury is detected in a larger proportion of patients who have non-ACS cardiopulmonary conditions than with previous, less sensitive assays. Therefore, other aspects of the clinical evaluation are critical to the practitioner's determination of the probability that the symptoms represent ACS. In addition, observation of a change in cardiac troponin concentration between serial samples is necessary for discriminating acute causes of myocardial injury from chronic elevation due to underlying structural heart disease, end-stage renal disease, or the rare presence of interfering antibodies. The diagnosis of MI is reserved for acute myocardial injury that is marked by a rising and/or falling pattern—with at least one value exceeding the 99th percentile reference limit—*and that is caused by ischemia*. Other nonischemic insults, such as myocarditis, may result in acute myocardial injury but should not be labeled MI **(Fig. 14-3)**.

Other laboratory assessments may include the D-dimer test to aid in exclusion of pulmonary embolism **(Chap. 279)**. Measurement of a B-type natriuretic peptide is useful when considered in conjunction with the clinical history and exam for the diagnosis of heart failure. B-type natriuretic peptides also provide prognostic information among patients with ACS and those with pulmonary embolism.

INTEGRATIVE DECISION-AIDS

Multiple clinical algorithms have been developed to aid in decision-making during the evaluation and disposition of patients with acute nontraumatic chest pain. Such decision-aids estimate either of two closely related but not identical probabilities: (1) the probability of a final diagnosis of ACS and (2) the probability of major cardiac events during short-term follow-up. Such decision-aids are used most commonly to identify patients with a low clinical probability of ACS who are candidates for discharge from the ED, with or without additional noninvasive testing. Goldman and Lee developed one of the first such decision-aids, using only the ECG and risk indicators—hypotension, pulmonary rales, and known ischemic heart disease—to categorize patients into four risk categories ranging from a <1% to a >16% probability of a major cardiovascular complication. Decision-aids used more commonly in current practice are shown in **Fig. 14-4**. Elements common across multiple risk stratification tools are (1) symptoms typical for ACS; (2) older age; (3) risk factors for or known atherosclerosis; (4) ischemic ECG abnormalities; and (5) elevated cardiac troponin level. Although, because of very low specificity, the overall diagnostic performance of such decision-aids is poor (area under the receiver operating curve, 0.55–0.65), in conjunction with the ECG and serial high-sensitivity cardiac troponin, they can help identify patients with a very low probability of ACS (e.g., <1%) or adverse cardiovascular events (<2% at 30 days). Clinical application of such integrated decision-aids or "accelerated diagnostic protocols" has been reported to achieve overall "miss rates" for ACS of <0.5% and may be useful for identifying patients who may be discharged without the need for additional cardiac testing.

Clinicians should differentiate between the algorithms discussed above and risk scores derived for stratification of prognosis (e.g., the TIMI and GRACE risk scores, **Chap. 275**) in patients *who already have an established diagnosis of ACS*. The latter risk scores were not designed to be used for

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

have determined the patient to remain eligible for further testing because of intermediate or undetermined risk, diagnostic coronary imaging with coronary computed tomographic (CT) angiography or functional testing, preferably with nuclear or echocardiographic imaging, is recommended

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

GRACE risk scores, **Chap. 275**) in patients *who already have an established diagnosis of ACS*. The latter risk scores were not designed to be used for *diagnostic* assessment.

CORONARY AND MYOCARDIAL STRESS IMAGING

Among patients for whom other life-threatening causes of chest pain have been reasonably excluded and serial biomarker and clinical assessment have determined the patient to remain eligible for further testing because of intermediate or undetermined risk, diagnostic coronary imaging with coronary computed tomographic (CT) angiography or functional testing, preferably with nuclear or echocardiographic imaging, is recommended. Patient characteristics (e.g., body habitus and renal function), prior cardiac testing, history of known coronary artery disease, existing contraindications for a given test modality, and patient preferences are considerations when choosing among these diagnostic tests (**Chaps. 241 and A9**).

CT Angiography

CT angiography has emerged as a preferred modality for the evaluation of patients with acute chest discomfort who are candidates for further testing after biomarker and clinical risk assessment (See Chap. 241). Coronary CT angiography is a sensitive technique for detection of obstructive coronary disease. CT appears to enhance the speed to disposition of patients with a low-intermediate probability for ACS, with its major strength being the negative predictive value of a finding of no significant stenosis or coronary plaque. In addition, contrast-enhanced CT can detect focal areas of myocardial injury in the acute setting. At the same time, CT angiography can exclude aortic dissection, pericardial effusion, and pulmonary embolism.

Stress Nuclear Perfusion Imaging or Stress Echocardiography

Functional testing with stress nuclear perfusion imaging and stress echocardiography are alternatives for the evaluation of patients with acute chest pain who are candidates for further testing and are preferred over coronary CT angiography in patients with known obstructive epicardial disease (See Chaps. 241 and A9). The selection of stress test modality may depend on institutional availability and expertise. Stress testing with myocardial imaging, either with nuclear perfusion imaging or echocardiography, offers superior diagnostic performance over exercise ECG. In patients selected for stress myocardial imaging who are able to exercise, exercise stress is preferred over pharmacologic testing. When available, positron emission tomography offers advantages of improved diagnostic performance and fewer nondiagnostic studies than single-photon emission CT.

Although functional testing is generally contraindicated in patients with ongoing chest pain, in selected patients with persistent pain and nondiagnostic ECG and biomarker data, resting myocardial perfusion images can be obtained; the absence of any perfusion abnormality substantially reduces the likelihood of coronary artery disease. In such a strategy, used in some centers, those with abnormal rest perfusion imaging, which cannot discriminate between old or new myocardial defects, usually must undergo additional evaluation.

EXERCISE ELECTROCARDIOGRAPHY

Exercise electrocardiography has historically been commonly employed for completion of risk stratification of patients who have undergone an initial evaluation that has not revealed a specific cause of chest discomfort and has identified a low risk of ACS. Early exercise testing is safe in patients without ongoing chest pain or high-risk findings and may assist in refining their prognostic assessment. However, for patients with chest pain for whom both cardiac troponin and clinical risk stratification have determined the patient to have *low* probability of ACS, there is insufficient evidence that stress testing or cardiac imaging improves their outcomes. This evolution in evidence supports a change from past practice in which outpatient stress testing within 72 hours was broadly used for patients with acute chest pain.

OTHER NONINVASIVE STUDIES

Other noninvasive imaging studies of the chest can be used selectively to provide additional diagnostic and prognostic information on patients with chest discomfort.

Echocardiography

Echocardiography (nonstress) is not necessarily routine in patients with chest discomfort. However, in patients with an uncertain diagnosis, particularly those with nondiagnostic ST elevation, ongoing symptoms, or hemodynamic instability, detection of abnormal regional wall motion provides evidence of possible ischemic dysfunction. Echocardiography is diagnostic in patients with mechanical complications of MI or in patients with pericardial tamponade. Transthoracic echocardiography is poorly sensitive for aortic dissection, although an intimal flap may sometimes be detected in the ascending aorta.

MRI

Cardiac magnetic resonance (CMR) imaging is an evolving, versatile technique for structural and functional evaluation of the heart and the

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

discomfort, CMR can be a useful modality for cardiac structural evaluation of patients with elevated cardiac troponin levels in the absence of definite coronary artery disease. CMR coronary angiography is in its early stages. MRI also permits highly accurate assessment for aortic dissection but is

MRI

Cardiac magnetic resonance (CMR) imaging is an evolving, versatile technique for structural and functional evaluation of the heart and the vasculature of the chest (See Chap. 241). CMR can be performed as a modality for pharmacologic stress perfusion imaging. Gadolinium-enhanced CMR can provide early detection of MI, defining areas of myocardial necrosis accurately, and can delineate patterns of myocardial disease that are often useful in discriminating ischemic from nonischemic myocardial injury. Although usually not practical for the urgent evaluation of acute chest discomfort, CMR can be a useful modality for cardiac structural evaluation of patients with elevated cardiac troponin levels in the absence of definite coronary artery disease. CMR coronary angiography is in its early stages. MRI also permits highly accurate assessment for aortic dissection but is infrequently used as the first test because CT and transesophageal echocardiography are usually more practical.

FIGURE 14-2

Association of chest pain characteristics with the probability of acute myocardial infarction (AMI). Note that a subsequent larger study showed a nonsignificant association with radiation to the right arm. *(Figure prepared from data in CJ Swap, JT Nagurney: JAMA 294:2623, 2005.)*

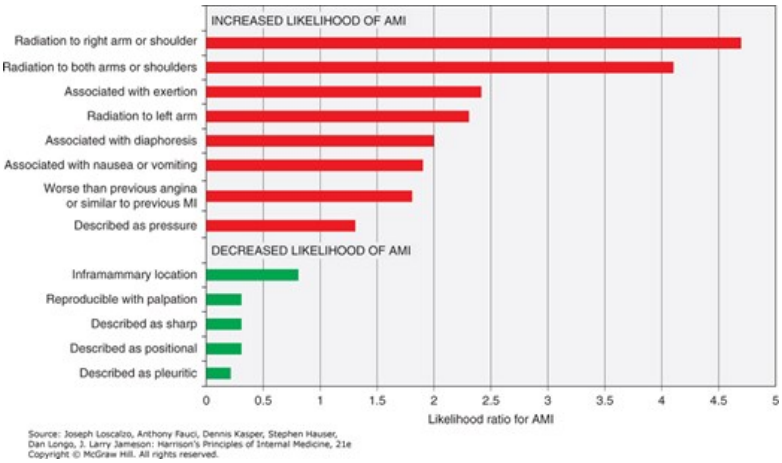


FIGURE 14-3

Clinical classification of patients with elevated cardiac troponin (cTn). MI, myocardial infarction.

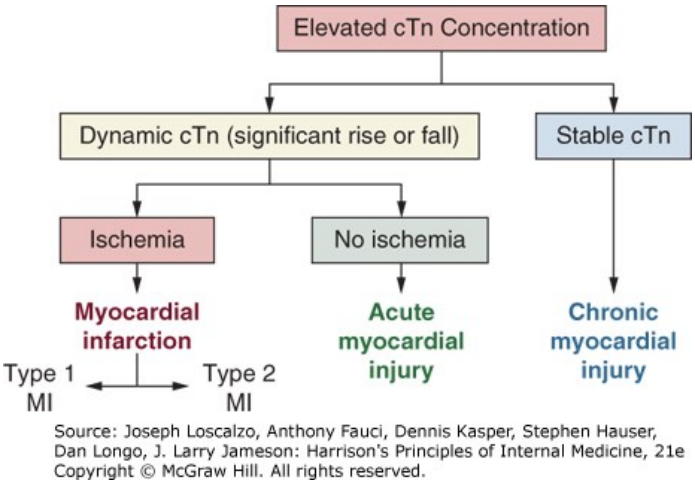


FIGURE 14-4

Examples of decision-aids used in conjunction with serial measurement of cardiac troponin (cTn) for evaluation of acute chest pain. The HEART score was modified by the authors in the presented study and omitting the assignment of 0, 1, or 2 points based on troponin. The negative predictive value (NPV) reported is for the composite endpoint of myocardial infarction (MI), cardiogenic shock, cardiac arrest, and all-cause

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

EDACS Score

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

negative predictive value (NPV) reported is for the composite endpoint of myocardial infarction (MI), cardiogenic shock, cardiac arrest, and all-cause mortality by 60 days. *Limit of quantification is the lowest analyte concentration that can be quantitatively detected with a total imprecision of ≤20%. CABG, coronary artery bypass graft; CAD, coronary artery disease; ECG, electrocardiogram; PCI, percutaneous coronary intervention. *(Figure prepared from data in Mark DG et al: J Am Coll Cardiol 13:606, 2018.)*

HEART Score (without cTn)		
History	Highly suspicious Moderately suspicious Slightly suspicious	2 1 0
ECG	Significant ST depression Nonspecific abnormality Normal	2 1 0
Age	≥65 y 45–<65 y <45 y	2 1 0
Risk factors	≥3 risk factors 1–2 risk factors None	2 1 0
	TOTAL	
	Low risk: 0–3 **Not low risk:** ≥4	

EDACS Score		
Age	86+ y 81–85 y 76–80 y Step down by 5-y increments 46–50 y 18–45 y	20 18 16 (–2) 4 2
Known CAD or risk factors	Known CAD (prior MI, PCI, or CABG) or ≥3 cardiac risk factors in patient aged ≤50 y	4
Sex	Male Female	6 0
Symptoms	Radiation to arm, shoulder, neck, or jaw Diaphoresis Pain with inspiration Reproduced by palpation	5 3 –4 –6
	TOTAL	
	Low risk: 0–15 **Not low risk:** ≥16	

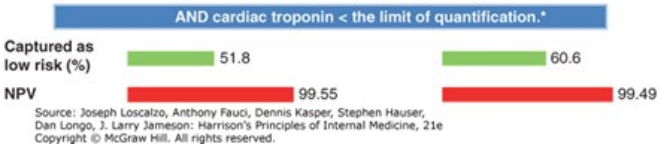


TABLE 14-2

Considerations in the Assessment of the Patient with Chest Discomfort

1. Could the chest discomfort be due to an acute, potentially life-threatening condition that warrants urgent evaluation and management?			
Unstable ischemic heart disease	Aortic dissection	Pneumothorax	Pulmonary embolism
2. If not, could the discomfort be due to a chronic condition likely to lead to serious complications?			
Stable angina	Aortic stenosis	Pulmonary hypertension	
3. If not, could the discomfort be due to an acute condition that warrants specific treatment?			
Pericarditis	Pneumonia/pleuritis	Herpes zoster	
4. If not, could the discomfort be due to another treatable chronic condition?			
Esophageal reflux		Cervical disk disease	
Esophageal spasm		Arthritis of the shoulder or spine	
Peptic ulcer disease		Costochondritis	
Gallbladder disease		Other musculoskeletal disorders	
Other gastrointestinal conditions		Anxiety state	

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 14-2

Considerations in the Assessment of the Patient with Chest Discomfort

1. Could the chest discomfort be due to an acute, potentially life-threatening condition that warrants urgent evaluation and management?			
Unstable ischemic heart disease	Aortic dissection	Pneumothorax	Pulmonary embolism
2. If not, could the discomfort be due to a chronic condition likely to lead to serious complications?			
Stable angina	Aortic stenosis	Pulmonary hypertension	
3. If not, could the discomfort be due to an acute condition that warrants specific treatment?			
Pericarditis	Pneumonia/pleuritis	Herpes zoster	
4. If not, could the discomfort be due to another treatable chronic condition?			
Esophageal reflux		Cervical disk disease	
Esophageal spasm		Arthritis of the shoulder or spine	
Peptic ulcer disease		Costochondritis	
Gallbladder disease		Other musculoskeletal disorders	
Other gastrointestinal conditions		Anxiety state	

Source: Developed by Dr. Thomas H. Lee for the 18th edition of *Harrison's Principles of Internal Medicine.*

CRITICAL PATHWAYS FOR ACUTE CHEST DISCOMFORT

Because of the challenges inherent in reliably identifying the small proportion of patients with serious causes of acute chest discomfort while not exposing the larger number of low-risk patients to unnecessary testing and extended ED or hospital evaluations, many medical centers have adopted critical pathways to expedite the assessment and management of patients with nontraumatic chest pain, often in dedicated chest pain units. Such pathways are generally aimed at (1) rapid identification, triage, and treatment of high-risk cardiopulmonary conditions (e.g., STEMI); (2) accurate identification of low-risk patients who can be safely observed in units with less intensive monitoring, undergo early noninvasive testing, or be discharged home; and (3) through more efficient and systematic accelerated diagnostic protocols, safe reduction in costs associated with overuse of testing and unnecessary hospitalizations. In some studies, provision of protocol-driven care in chest pain units has decreased costs and overall duration of hospital evaluation with no detectable excess of adverse clinical outcomes.

OUTPATIENT EVALUATION OF CHEST DISCOMFORT

Chest pain is common in outpatient practice, with a lifetime prevalence of 20–40% in the general population. More than 25% of patients with MI have had a related visit with a primary care physician in the previous month. The diagnostic principles are the same as in the ED. However, the pretest probability of an acute cardiopulmonary cause is significantly lower. Therefore, testing paradigms are less intense, with an emphasis on the history, physical examination, and ECG. Moreover, decision-aids developed for settings with a high prevalence of significant cardiopulmonary disease have lower positive predictive value when applied in the practitioner's office. However, in general, if the level of clinical suspicion of ACS is sufficiently high to consider troponin testing, the patient should be referred to the ED for evaluation.

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility
Amsterdam EA et al: Testing of low-risk patients presenting to the emergency department with chest pain: A scientific statement from the American Heart Association. Circulation 122:1756. 2010. [PubMed: 20660809]

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

to consider troponin testing, the patient should be referred to the ED for evaluation.

FURTHER READING

Amsterdam EA et al: Testing of low-risk patients presenting to the emergency department with chest pain: A scientific statement from the American Heart Association. Circulation 122:1756, 2010. [PubMed: 20660809]

Chapman AR et al: Association of high-sensitivity cardiac troponin I concentration with cardiac outcomes in patients with suspected acute coronary syndrome. JAMA 318:1913, 2017. [PubMed: 29127948]

Fanaroff AC et al: Does this patient with chest pain have acute coronary syndrome? JAMA 314:1955, 2015. [PubMed: 26547467]

Hsia RY et al: A national study of the prevalence of life-threatening diagnoses in patients with chest pain. JAMA Intern Med 176:1029, 2016. [PubMed: 27295579]

Mahler SA et al: Safely identifying emergency department patients with acute chest pain for early discharge: HEART pathway accelerated diagnostic protocol. Circulation 138:2456, 2018. [PubMed: 30571347]

Downloaded 2022-1-3 5:43 P Your IP is 141.216.78.40
Chapter 14: Chest Discomfort, David A. Morrow
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 15: Abdominal Pain

Danny O. Jacobs

INTRODUCTION

Correctly diagnosing acute abdominal pain can be quite challenging. Few clinical situations require greater judgment, because the most catastrophic of events may be forecast by the subtlest of symptoms and signs. In every instance, the clinician must distinguish those conditions that require urgent intervention from those that do not and can best be managed nonoperatively. A meticulously executed, detailed history and physical examination are critically important for focusing the differential diagnosis and allowing the diagnostic evaluation to proceed expeditiously **(Table 15-1)**.

TABLE 15-1

Some Key Components of the Patient's History

Age
Time and mode of onset of the pain
Pain characteristics
Duration of symptoms
Location of pain and sites of radiation
Associated symptoms and their relationship to the pain
Nausea, emesis, and anorexia
Diarrhea, constipation, or other changes in bowel habits
Menstrual history

The etiologic classification in **Table 15-2**, although not complete, provides a useful framework for evaluating patients with abdominal pain.

TABLE 15-2

Some Important Causes of Abdominal Pain

Pain Originating in the Abdomen	
Parietal peritoneal inflammation	Vascular disturbances
Bacterial contamination	Embolism or thrombosis
Perforated appendix or other perforated viscus	Vascular rupture
Pelvic inflammatory disease	Pressure or torsional occlusion
Chemical irritation	Sickle cell anemia
Perforated ulcer	Abdominal wall
Mechanical obstruction of hollow viscera	Distension of visceral surfaces, e.g., by hemorrhage
Obstruction of the small or large intestine	Hepatic or renal capsules

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 15: Abdominal Pain, Danny O. Jacobs
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Parietal peritoneal inflammation	Vascular disturbances
Chemical irritation	Sickle cell anemia
Perforated ulcer	Abdominal wall
Pancreatitis	Distortion or traction of mesentery
Mittelschmerz	Trauma or infection of muscles
Mechanical obstruction of hollow viscera	Distension of visceral surfaces, e.g., by hemorrhage
Obstruction of the small or large intestine	Hepatic or renal capsules
Obstruction of the biliary tree	Inflammation
Obstruction of the ureter	Appendicitis
	Typhoid fever
	Neutropenic enterocolitis or “typhlitis”
Pain Referred from Extraabdominal Source	
Cardiothoracic	Pleurodynia
Acute myocardial infarction	Pneumothorax
Myocarditis, endocarditis, pericarditis	Empyema
Congestive heart failure	Esophageal disease, including spasm, rupture, or inflammation
Pneumonia (especially lower lobes)	Genitalia
Pulmonary embolus	Torsion of the testis
Metabolic Causes	
Diabetes	Acute adrenal insufficiency
Uremia	Familial Mediterranean fever
Hyperlipidemia	Porphyria
Hyperparathyroidism	C1 esterase inhibitor deficiency (angioneurotic edema)
Neurologic/Psychiatric Causes	
Herpes zoster	Spinal cord or nerve root compression
Tabes dorsalis	Functional disorders
Causalgia	Psychiatric disorders
Radiculitis from infection or arthritis	
Toxic Causes	
Lead poisoning	
Insect or animal envenomation	
Black widow spider bites	
Snake bites	
Uncertain Mechanisms	
Narcotic withdrawal	
Heat stroke	

Any patient with abdominal pain of recent onset requires an early and thorough evaluation. The most common causes of abdominal pain on admission are nonspecific abdominal pain, acute appendicitis, pain of urologic origin, and intestinal obstruction. A diagnosis of “acute or surgical abdomen” is not acceptable because of its often misleading and erroneous connotations. Most patients who present with acute abdominal pain will have self-limited disease processes. However, it is important to remember that pain severity does not necessarily correlate with the severity of the underlying condition. And, the presence or absence of various degrees of “hunger” is unreliable as a sole indicator of the severity of intraabdominal

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 15: Abdominal Pain, Danny O. Jacobs
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

SOME MECHANISMS OF PAIN ORIGINATING IN THE ABDOMEN

Inflammation of the Parietal Peritoneum

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

underlying condition. And, the presence or absence of various degrees of “hunger” is unreliable as a sole indicator of the severity of intraabdominal disease. The most obvious of “acute abdomens” may not require operative intervention, and the mildest of abdominal pains may herald an urgently correctable disease.

SOME MECHANISMS OF PAIN ORIGINATING IN THE ABDOMEN

Inflammation of the Parietal Peritoneum

The pain of parietal peritoneal inflammation is steady and aching in character and is located directly over the inflamed area, its exact reference being possible because it is transmitted by somatic nerves supplying the parietal peritoneum. The intensity of the pain is dependent on the type and amount of material to which the peritoneal surfaces are exposed in a given time period. For example, the sudden release of a small quantity of *sterile* acidic gastric juice into the peritoneal cavity causes much more pain than the same amount of grossly contaminated neutral feces. Enzymatically active pancreatic juice incites more pain and inflammation than does the same amount of sterile bile containing no potent enzymes. Blood is normally only a mild irritant, and the response to urine is also typically bland, so exposure of blood and urine to the peritoneal cavity may go unnoticed unless it is sudden and massive. Bacterial contamination, such as may occur with pelvic inflammatory disease or perforated distal intestine, causes low-intensity pain until multiplication causes significant amounts of inflammatory mediators to be released. Patients with perforated upper gastrointestinal ulcers may present entirely differently depending on how quickly gastric juices enter the peritoneal cavity and their pH. Thus, the rate at which any inflammatory material irritates the peritoneum is important.

The pain of peritoneal inflammation is invariably accentuated by pressure or changes in tension of the peritoneum, whether produced by palpation or by movement such as with coughing or sneezing. The patient with peritonitis characteristically lies quietly in bed, preferring to avoid motion, in contrast to the patient with colic, who may be thrashing in discomfort.

Another characteristic feature of peritoneal irritation is tonic reflex spasm of the abdominal musculature, localized to the involved body segment. Its intensity depends on the integrity of the nervous system, the location of the inflammatory process, and the rate at which it develops. Spasm over a perforated retrocecal appendix or perforation into the lesser peritoneal sac may be minimal or absent because of the protective effect of overlying viscera. Catastrophic abdominal emergencies may be associated with minimal or no detectable pain or muscle spasm in obtunded, seriously ill, debilitated, immunosuppressed, or psychotic patients. A slowly developing process also often greatly attenuates the degree of muscle spasm.

Obstruction of Hollow Viscera

Intraluminal obstruction classically elicits intermittent or colicky abdominal pain that is not as well localized as the pain of parietal peritoneal irritation. However, the absence of cramping discomfort can be misleading because distention of a hollow viscus may also produce steady pain with only rare paroxysms.

Small-bowel obstruction often presents as poorly localized, intermittent periumbilical or supraumbilical pain. As the intestine progressively dilates and loses muscular tone, the colicky nature of the pain may diminish. With superimposed strangulating obstruction, pain may spread to the lower lumbar region if there is traction on the root of the mesentery. The colicky pain of colonic obstruction is of lesser intensity, is commonly located in the infraumbilical area, and may often radiate to the lumbar region.

Sudden distention of the biliary tree produces a steady rather than colicky type of pain; hence, the term *biliary colic* is misleading. Acute distention of the gallbladder typically causes pain in the right upper quadrant with radiation to the right posterior region of the thorax or to the tip of the right scapula, but discomfort is also not uncommonly found near the midline. Distention of the common bile duct often causes epigastric pain that may radiate to the upper lumbar region. Considerable variation is common, however, so that differentiation between gallbladder or common ductal disease may be impossible.

Gradual dilatation of the biliary tree, as can occur with carcinoma of the head of the pancreas, may cause no pain or only a mild aching sensation in the epigastrium or right upper quadrant. The pain of distention of the pancreatic ducts is similar to that described for distention of the common bile duct but, in addition, is very frequently accentuated by recumbency and relieved by the upright position.

Obstruction of the urinary bladder usually causes dull, low-intensity pain in the suprapubic region. Restlessness, without specific complaint of pain, may be the only sign of a distended bladder in an obtunded patient. In contrast, acute obstruction of the intravesicular portion of the ureter is characterized by severe suprapubic and flank pain that radiates to the penis, scrotum, or inner aspect of the upper thigh. Obstruction of the ureteropelvic junction manifests as pain near the costovertebral angle, whereas obstruction of the remainder of the ureter is associated with flank

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 15: Abdominal Pain, Danny O. Jacobs
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

A frequent misconception is that pain due to intraabdominal vascular disturbances is sudden and catastrophic in nature. Certain disease processes,

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

characterized by severe suprapubic and flank pain that radiates to the penis, scrotum, or inner aspect of the upper thigh. Obstruction of the ureteropelvic junction manifests as pain near the costovertebral angle, whereas obstruction of the remainder of the ureter is associated with flank pain that often extends into the same side of the abdomen.

Vascular Disturbances

A frequent misconception is that pain due to intraabdominal vascular disturbances is sudden and catastrophic in nature. Certain disease processes, such as embolism or thrombosis of the superior mesenteric artery or impending rupture of an abdominal aortic aneurysm, can certainly be associated with diffuse, severe pain. Yet, just as frequently, the patient with occlusion of the superior mesenteric artery only has mild continuous or cramping diffuse pain for 2 or 3 days before vascular collapse or findings of peritoneal inflammation appear. The early, seemingly insignificant discomfort is caused by hyperperistalsis rather than peritoneal inflammation. Indeed, absence of tenderness and rigidity in the presence of continuous, diffuse pain (e.g., “pain out of proportion to physical findings”) in a patient likely to have vascular disease is quite characteristic of occlusion of the superior mesenteric artery. Abdominal pain with radiation to the sacral region, flank, or genitalia should always signal the possible presence of a rupturing abdominal aortic aneurysm. This pain may persist over a period of several days before rupture and collapse occur.

Abdominal Wall

Pain arising from the abdominal wall is usually constant and aching. Movement, prolonged standing, and pressure accentuate the discomfort and associated muscle spasm. In the relatively rare case of hematoma of the rectus sheath, now most frequently encountered in association with anticoagulant therapy, a mass may be present in the lower quadrants of the abdomen. Simultaneous involvement of muscles in other parts of the body usually serves to differentiate myositis of the abdominal wall from other processes that might cause pain in the same region.

REFERRED PAIN IN ABDOMINAL DISEASE

Pain referred to the abdomen from the thorax, spine, or genitalia may present a diagnostic challenge because diseases of the upper part of the abdominal cavity such as acute cholecystitis or perforated ulcer may be associated with intrathoracic complications. A most important, yet often forgotten, dictum is that the possibility of intrathoracic disease must be considered in every patient with abdominal pain, especially if the pain is in the upper abdomen.

Systematic questioning and examination directed toward detecting myocardial or pulmonary infarction, pneumonia, pericarditis, or esophageal disease (the intrathoracic diseases that most often masquerade as abdominal emergencies) will often provide sufficient clues to establish the proper diagnosis. Diaphragmatic pleuritis resulting from pneumonia or pulmonary infarction may cause pain in the right upper quadrant and pain in the supraclavicular area, the latter radiation to be distinguished from the referred subscapular pain caused by acute distention of the extrahepatic biliary tree. The ultimate decision as to the origin of abdominal pain may require deliberate and planned observation over a period of several hours, during which repeated questioning and examination will provide the diagnosis or suggest the appropriate studies.

Referred pain of thoracic origin is often accompanied by splinting of the involved hemithorax with respiratory lag and a decrease in excursion more marked than that seen in the presence of intraabdominal disease. In addition, apparent abdominal muscle spasm caused by referred pain will diminish during the inspiratory phase of respiration, whereas it persists throughout both respiratory phases if it is of abdominal origin. Palpation over the area of referred pain in the abdomen also does not usually accentuate the pain and, in many instances, actually seems to relieve it.

Thoracic disease and abdominal disease frequently coexist and may be difficult or impossible to differentiate. For example, the patient with known biliary tract disease often has epigastric pain during myocardial infarction, or biliary colic may be referred to the precordium or left shoulder in a patient who has suffered previously from angina pectoris. **For an explanation of the radiation of pain to a previously diseased area, see Chap. 13.**

Referred pain from the spine, which usually involves compression or irritation of nerve roots, is characteristically intensified by certain motions such as cough, sneeze, or strain and is associated with hyperesthesia over the involved dermatomes. Pain referred to the abdomen from the testes or seminal vesicles is generally accentuated by the slightest pressure on either of these organs. The abdominal discomfort experienced is of dull, aching character and is poorly localized.

METABOLIC ABDOMINAL CRISES

Pain of metabolic origin may simulate almost any other type of intraabdominal disease. Several mechanisms may be at work. In certain instances, such pain. Whenever the cause of abdominal pain is obscure, a metabolic origin always must be considered. Abdominal pain is also the hallmark of familial Mediterranean fever **(Chap. 369)**.

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 15: Abdominal Pain, Danny O. Jacobs
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Pain of metabolic origin may simulate almost any other type of intraabdominal disease. Several mechanisms may be at work. In certain instances, such as hyperlipidemia, the metabolic disease itself may be accompanied by an intraabdominal process such as pancreatitis, which can lead to unnecessary laparotomy unless recognized. C1 esterase deficiency associated with angioneurotic edema is often associated with episodes of severe abdominal pain. Whenever the cause of abdominal pain is obscure, a metabolic origin always must be considered. Abdominal pain is also the hallmark of familial Mediterranean fever **(Chap. 369)**.

The pain of porphyria and of lead colic is usually difficult to distinguish from that of intestinal obstruction, because severe hyperperistalsis is a prominent feature of both. The pain of uremia or diabetes is nonspecific, and the pain and tenderness frequently shift in location and intensity. Diabetic acidosis may be precipitated by acute appendicitis or intestinal obstruction, so if prompt resolution of the abdominal pain does not result from correction of the metabolic abnormalities, an underlying organic problem should be suspected. Black widow spider bites produce intense pain and rigidity of the abdominal muscles and back, an area infrequently involved in intraabdominal disease.

IMMUNOCOMPROMISE

Evaluating and diagnosing causes of abdominal pain in immunosuppressed or otherwise immunocompromised patients is very difficult. This includes those who have undergone organ transplantation; who are receiving immunosuppressive treatments for autoimmune diseases, chemotherapy, or glucocorticoids; who have AIDS; and who are very old. In these circumstances, normal physiologic responses may be absent or masked. In addition, unusual infections may cause abdominal pain where the etiologic agents include cytomegalovirus, mycobacteria, protozoa, and fungi. These pathogens may affect all gastrointestinal organs, including the gallbladder, liver, and pancreas, as well as the gastrointestinal tract, causing occult or overtly symptomatic perforations of the latter. Splenic abscesses due to *Candida* or *Salmonella* infection should also be considered, especially when evaluating patients with left upper quadrant or left flank pain. Acalculous cholecystitis may be observed in immunocompromised patients or those with AIDS, where it is often associated with cryptosporidiosis or cytomegalovirus infection.

Neutropenic enterocolitis (typhlitis) is often identified as a cause of abdominal pain and fever in some patients with bone marrow suppression due to chemotherapy. Acute graft-versus-host disease should be considered in this circumstance. Optimal management of these patients requires meticulous follow-up including serial examinations to assess the need for more surgical intervention, for example, to address perforation.

NEUROGENIC CAUSES

Diseases that injure sensory nerves may cause causalgic pain. This pain has a burning character and is usually limited to the distribution of a given peripheral nerve. Stimuli that are normally not painful such as touch or a change in temperature may be causalgic and are often present even at rest. The demonstration of irregularly spaced cutaneous “pain spots” may be the only indication that an old nerve injury exists. Even though the pain may be precipitated by gentle palpation, rigidity of the abdominal muscles is absent, and the respirations are not usually disturbed. Distention of the abdomen is uncommon, and the pain has no relationship to food intake.

Pain arising from spinal nerves or roots comes and goes suddenly and is of a lancinating type **(Chap. 17)**. It may be caused by herpes zoster, impingement by arthritis, tumors, a herniated nucleus pulposus, diabetes, or syphilis. It is not associated with food intake, abdominal distention, or changes in respiration. Severe muscle spasms, when present, may be relieved by, but are certainly not accentuated by, abdominal palpation. The pain is made worse by movement of the spine and is usually confined to a few dermatomes. Hyperesthesia is very common.

Pain due to functional causes conforms to none of the aforementioned patterns. Mechanisms of disease are not clearly established. Irritable bowel syndrome (IBS) is a functional gastrointestinal disorder characterized by abdominal pain and altered bowel habits. The diagnosis is made on the basis of clinical criteria **(Chap. 327)** and after exclusion of demonstrable structural abnormalities. The episodes of abdominal pain may be brought on by stress, and the pain varies considerably in type and location. Nausea and vomiting are rare. Localized tenderness and muscle spasm are inconsistent or absent. The causes of IBS or related functional disorders are not yet fully understood.

APPROACH TO THE PATIENT

Abdominal Pain

Few abdominal conditions require such urgent operative intervention that an orderly approach needs to be abandoned, no matter how ill the patient is. Only patients with exsanguinating intraabdominal hemorrhage (e.g., ruptured aneurysm) must be rushed to the operating room

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 15: Abdominal Pain, Danny O. Jacobs
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

contraindications to operation when massive intraabdominal hemorrhage is present. Fortunately, this situation is relatively rare. This statement does not necessarily apply to patients with intraluminal gastrointestinal hemorrhage, who can often be managed by other means **(Chap. 48)**. In

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

patient is. Only patients with exsanguinating intraabdominal hemorrhage (e.g., ruptured aneurysm) must be rushed to the operating room immediately, but in such instances, only a few minutes are required to assess the critical nature of the problem. Under these circumstances, all obstacles must be swept aside, adequate venous access for fluid replacement obtained, and the operation begun. Unfortunately, many of these patients may die in the radiology department or the emergency room while awaiting unnecessary examinations. *There are no absolute contraindications to operation when massive intraabdominal hemorrhage is present.* Fortunately, this situation is relatively rare. This statement does not necessarily apply to patients with intraluminal gastrointestinal hemorrhage, who can often be managed by other means **(Chap. 48)**. In these patients, obtaining a *detailed history when possible* can be extremely helpful even though it can be laborious and time-consuming. Decision-making regarding next steps is facilitated and a reasonably accurate diagnosis can be made before any further diagnostic testing is undertaken.

In cases of *acute* abdominal pain, a diagnosis can be readily established in most instances, whereas success is not so frequent in patients with *chronic* pain. IBS is one of the most common causes of abdominal pain and must always be kept in mind **(Chap. 327)**. The location of the pain can assist in narrowing the differential diagnosis **(Table 15-3)**; however, the *chronological sequence of events* in the patient's history is often more important than the pain's location. Careful attention should be paid to the extraabdominal regions. Narcotics or analgesics should *not* be withheld until a definitive diagnosis or a definitive plan has been formulated; obfuscation of the diagnosis by adequate analgesia is unlikely.

An accurate menstrual history in a female patient is essential. It is important to remember that normal anatomic relationships can be significantly altered by the gravid uterus. Abdominal and pelvic pain may occur during pregnancy due to conditions that do not require operation. Lastly, some otherwise noteworthy laboratory values (e.g., leukocytosis) may represent the normal physiologic changes of pregnancy.

In the examination, simple critical inspection of the patient, for example, of facies, position in bed, and respiratory activity, provides valuable clues. The amount of information to be gleaned is directly proportional to the *gentleness* and thoroughness of the examiner. Once a patient with peritoneal inflammation has been examined brusquely, accurate assessment by the next examiner becomes almost impossible. Eliciting rebound tenderness by sudden release of a deeply palpating hand in a patient with suspected peritonitis is cruel and unnecessary. The same information can be obtained by gentle percussion of the abdomen (rebound tenderness on a miniature scale), a maneuver that can be far more precise and localizing. Asking the patient to cough will elicit true rebound tenderness without the need for placing a hand on the abdomen. Furthermore, the forceful demonstration of rebound tenderness will startle and induce protective spasm in a nervous or worried patient in whom true rebound tenderness is not present. A palpable gallbladder will be missed if palpation is so aggressive that voluntary muscle spasm becomes superimposed on involuntary muscular rigidity.

As with history taking, sufficient time should be spent in the examination. Abdominal signs may be minimal but, nevertheless, if accompanied by consistent symptoms, may be exceptionally meaningful. Abdominal signs may be virtually or totally absent in cases of pelvic peritonitis, so careful *pelvic and rectal examinations are mandatory in every patient with abdominal pain.* Tenderness on pelvic or rectal examination in the absence of other abdominal signs can be caused by operative indications such as perforated appendicitis, diverticulitis, twisted ovarian cyst, and many others. Much attention has been paid to the presence or absence of peristaltic sounds, their quality, and their frequency. Auscultation of the abdomen is one of the least revealing aspects of the physical examination of a patient with abdominal pain. Catastrophes such as a strangulating small-intestinal obstruction or perforated appendicitis may occur in the presence of normal peristaltic sounds. Conversely, when the proximal part of the intestine above obstruction becomes markedly distended and edematous, peristaltic sounds may lose the characteristics of borborygmi and become weak or absent, even when peritonitis is not present. It is usually the severe chemical peritonitis of sudden onset that is associated with the truly silent abdomen.

Laboratory examinations may be valuable in assessing the patient with abdominal pain, yet, with few exceptions, they rarely establish a diagnosis. Leukocytosis should never be the single deciding factor as to whether or not operation is indicated. A white blood cell count >20,000/μL may be observed with perforation of a viscus, but pancreatitis, acute cholecystitis, pelvic inflammatory disease, and intestinal infarction may also be associated with marked leukocytosis. A normal white blood cell count is not rare in cases of perforation of abdominal viscera. A diagnosis of anemia may be more helpful than the white blood cell count, especially when combined with the history.

The urinalysis may reveal the state of hydration or rule out severe renal disease, diabetes, or urinary infection. Blood urea nitrogen, glucose, and serum bilirubin levels and liver function tests may be helpful. Serum amylase levels may be increased by many diseases other than pancreatitis, for example, perforated ulcer, strangulating intestinal obstruction, and acute cholecystitis; thus, elevations of serum amylase do not rule in or rule out the need for an operation.

Plain and upright or lateral decubitus radiographs of the abdomen have limited utility and may be unnecessary in some patients who have substantial evidence of some diseases such as acute appendicitis or strangulated external hernia. Where the indications for surgical or medical

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Very rarely, barium or water-soluble contrast study of the upper part of the gastrointestinal tract is an appropriate radiographic investigation and may demonstrate partial intestinal obstruction that may elude diagnosis by other means. If there is any question of obstruction of the colon, oral

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Plain and upright or lateral decubitus radiographs of the abdomen have limited utility and may be unnecessary in some patients who have substantial evidence of some diseases such as acute appendicitis or strangulated external hernia. Where the indications for surgical or medical intervention are not clear, low-dose computed tomography is preferred to abdominal radiography when evaluating nontraumatic acute abdominal pain.

Very rarely, barium or water-soluble contrast study of the upper part of the gastrointestinal tract is an appropriate radiographic investigation and may demonstrate partial intestinal obstruction that may elude diagnosis by other means. If there is any question of obstruction of the colon, oral administration of barium sulfate should be avoided. On the other hand, in cases of suspected colonic obstruction (without perforation), a contrast enema may be diagnostic.

In the absence of trauma, peritoneal lavage has been replaced as a diagnostic tool by CT scanning and laparoscopy. Ultrasonography has proved to be useful in detecting an enlarged gallbladder or pancreas, the presence of gallstones, an enlarged ovary, or a tubal pregnancy. Laparoscopy is especially helpful in diagnosing pelvic conditions, such as ovarian cysts, tubal pregnancies, salpingitis, acute appendicitis, and other disease processes. Laparoscopy has a particular advantage over imaging in that the underlying etiologic condition can often be definitively addressed.

Radioisotopic hepatobiliary iminodiacetic acid scans (HIDAs) may help differentiate acute cholecystitis or biliary colic from acute pancreatitis. A CT scan may demonstrate an enlarged pancreas, ruptured spleen, or thickened colonic or appendiceal wall and streaking of the mesocolon or mesoappendix characteristic of diverticulitis or appendicitis.

Sometimes, even under the best circumstances with all available aids and with the greatest of clinical skill, a definitive diagnosis cannot be established at the time of the initial examination. And, in some cases, operation may be indicated based on clinical grounds alone. Should that decision be questionable, watchful waiting with repeated questioning and examination will often elucidate the true nature of the illness and indicate the proper course of action.

TABLE 15-3

Differential Diagnoses of Abdominal Pain by Location

Right Upper Quadrant	**Epigastric**	**Left Upper Quadrant**
Cholecystitis Cholangitis Pancreatitis Pneumonia/empyema Pleurisy/pleurodynia Subdiaphragmatic abscess Hepatitis Budd-Chiari syndrome	Peptic ulcer disease Gastritis GERD Pancreatitis Myocardial infarction Pericarditis Ruptured aortic aneurysm Esophagitis	Splenic infarct Splenic rupture Splenic abscess Gastritis Gastric ulcer Pancreatitis Subdiaphragmatic abscess
Right Lower Quadrant	**Periumbilical**	**Left Lower Quadrant**
Appendicitis Salpingitis Inguinal hernia Ectopic pregnancy Nephrolithiasis Inflammatory bowel disease Mesenteric lymphadenitis Typhlitis	Early appendicitis Gastroenteritis Bowel obstruction Ruptured aortic aneurysm	Diverticulitis Salpingitis Inguinal hernia Ectopic pregnancy Nephrolithiasis Irritable bowel syndrome Inflammatory bowel disease
Diffuse Nonlocalized Pain		
Irritable bowel syndrome Peritonitis	Psychiatric disease	

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 15: Abdominal Pain, Danny O. Jacobs
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 15-3

Differential Diagnoses of Abdominal Pain by Location

Right Upper Quadrant	**Epigastric**	**Left Upper Quadrant**
Cholecystitis Cholangitis Pancreatitis Pneumonia/empyema Pleurisy/pleurodynia Subdiaphragmatic abscess Hepatitis Budd-Chiari syndrome	Peptic ulcer disease Gastritis GERD Pancreatitis Myocardial infarction Pericarditis Ruptured aortic aneurysm Esophagitis	Splenic infarct Splenic rupture Splenic abscess Gastritis Gastric ulcer Pancreatitis Subdiaphragmatic abscess
Right Lower Quadrant	**Periumbilical**	**Left Lower Quadrant**
Appendicitis Salpingitis Inguinal hernia Ectopic pregnancy Nephrolithiasis Inflammatory bowel disease Mesenteric lymphadenitis Typhlitis	Early appendicitis Gastroenteritis Bowel obstruction Ruptured aortic aneurysm	Diverticulitis Salpingitis Inguinal hernia Ectopic pregnancy Nephrolithiasis Irritable bowel syndrome Inflammatory bowel disease
Diffuse Nonlocalized Pain		
Gastroenteritis Mesenteric ischemia Bowel obstruction Irritable bowel syndrome Peritonitis Diabetes	Malaria Familial Mediterranean fever Metabolic diseases Psychiatric disease	

Abbreviation: GERD, gastroesophageal reflux disease.

ACKNOWLEDGMENT

The author gratefully acknowledges the enormous contribution to this chapter and the approach it espouses of William Silen, who wrote this chapter for many editions.

FURTHER READING

Bhangu A et al: Acute appendicitis: Modern understanding of pathogenesis, diagnosis and management. Lancet 386:1278, 2015. [PubMed: 26460662]

Cartwright SL, Knudson MP: Diagnostic imaging of acute abdominal pain in adults. Am Fam Phys 91:452, 2015.

Huckins DS et al: Diagnostic performance of a biomarker panel as a negative predictor for acute appendicitis in acute emergency department patients

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 15: Abdominal Pain, Danny O. Jacobs
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Phillips MT: Clinical yield of computed tomography scans in the emergency department for abdominal pain. J Invest Med 64:542, 2016.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Huckins DS et al: Diagnostic performance of a biomarker panel as a negative predictor for acute appendicitis in acute emergency department patients with abdominal pain. Am J Emerg Med 35:418, 2017. [PubMed: 27876541]

Nayor J et al: Tracing the cause of abdominal pain. N Engl J Med 375:e8, 2016. [PubMed: 27509123]

Phillips MT: Clinical yield of computed tomography scans in the emergency department for abdominal pain. J Invest Med 64:542, 2016.

Silen W, Cope Z: *Cope's Early Diagnosis of the Acute Abdomen*, 22nd ed. New York, Oxford University Press, 2010.

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 15: Abdominal Pain, Danny O. Jacobs
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

Chapter 16: Headache

Peter J. Goadsby

INTRODUCTION

Headache is among the most common reasons patients seek medical attention and is responsible, on a global basis, for more disability than any other neurologic problem. Diagnosis and management are based on a careful clinical approach augmented by an understanding of the anatomy, physiology, and pharmacology of the nervous system pathways mediating the various headache syndromes. This chapter will focus on the general approach to a patient with headache; migraine and other primary headache disorders are discussed in **Chap. 430**.

GENERAL PRINCIPLES

A classification system developed by the International Headache Society (*www.ihs-headache.org/en/resources/guidelines/*) characterizes headache as primary or secondary **(Table 16-1)**. *Primary headaches* are those in which headache and its associated features are the disorder itself, whereas *secondary headaches* are those caused by exogenous disorders (Headache Classification Committee of the International Headache Society, 2018). Primary headache often results in considerable disability and a decrease in the patient's quality of life. Mild secondary headache, such as that seen in association with upper respiratory tract infections, is common but rarely worrisome. Life-threatening headache is relatively uncommon, but vigilance is required in order to recognize and appropriately treat such patients.

TABLE 16-1

Common Causes of Headache

PRIMARY HEADACHE		SECONDARY HEADACHE	
TYPE	**%**	**TYPE**	**%**
Tension-type	69	Systemic infection	63
Migraine	16	Head injury	4
Idiopathic stabbing	2	Vascular disorders	1
Exertional	1	Subarachnoid hemorrhage	<1
Cluster	0.1	Brain tumor	0.1

Source: After J Olesen et al: *The Headaches*. Philadelphia, Lippincott Williams & Wilkins, 2005.

ANATOMY AND PHYSIOLOGY OF HEADACHE

Pain usually occurs when peripheral nociceptors are stimulated in response to tissue injury, visceral distension, or other factors **(Chap. 13)**. In such situations, pain perception is a normal physiologic response mediated by a healthy nervous system. Pain can also result when pain-producing pathways of the peripheral or central nervous system (CNS) are damaged or activated inappropriately. Headache may originate from either or both mechanisms. Relatively few cranial structures are pain producing; these include the scalp, meningeal arteries, dural sinuses, falx cerebri, and proximal

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

- The large intracranial vessels and dura mater, and the peripheral terminals of the trigeminal nerve that innervate these structures

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

pathways of the peripheral or central nervous system (CNS) are damaged or activated inappropriately. Headache may originate from either or both mechanisms. Relatively few cranial structures are pain producing; these include the scalp, meningeal arteries, dural sinuses, falx cerebri, and proximal segments of the large pial arteries. The ventricular ependyma, choroid plexus, pial veins, and much of the brain parenchyma are not pain producing.

The key structures involved in primary headache are the following:

- The large intracranial vessels and dura mater, and the peripheral terminals of the trigeminal nerve that innervate these structures
- The caudal portion of the trigeminal nucleus, which extends into the dorsal horns of the upper cervical spinal cord and receives input from the first and second cervical nerve roots (the trigeminocervical complex)
- Rostral pain-processing regions, such as the ventroposteromedial thalamus and the cortex
- The pain-modulatory systems in the brain that modulate input from the trigeminal nociceptors at all levels of the pain-processing pathways and influence vegetative functions, such as the hypothalamus and brainstem

The *trigeminovascular system* innervates the large intracranial vessels and dura mater via the trigeminal nerve. Cranial autonomic symptoms, such as lacrimation, conjunctival injection, nasal congestion, rhinorrhea, periorbital swelling, aural fullness, and ptosis, are prominent in the trigeminal autonomic cephalalgias (TACs), including cluster headache and paroxysmal hemicrania, and may also be seen in migraine, even in children. These autonomic symptoms reflect activation of cranial parasympathetic pathways, and functional imaging studies indicate that vascular changes in migraine and cluster headache, when present, are similarly driven by these cranial autonomic systems. Thus, they are secondary, and not causative, events in the headache cascade. Moreover, they can often be mistaken for symptoms or signs of cranial sinus inflammation, which is then overdiagnosed and inappropriately managed. Migraine and other primary headache types are not “vascular headaches”; these disorders do not reliably manifest vascular changes, and treatment outcomes cannot be predicted by vascular effects. Migraine is a brain disorder and is best understood and managed as such.

CLINICAL EVALUATION OF ACUTE, NEW-ONSET HEADACHE

The patient who presents with a new, severe headache has a differential diagnosis that is quite different from the patient with recurrent headaches over many years. In new-onset and severe headache, the probability of finding a potentially serious cause is considerably greater than in recurrent headache. Patients with recent onset of pain require prompt evaluation and appropriate treatment. Serious causes to be considered include meningitis, subarachnoid hemorrhage, epidural or subdural hematoma, glaucoma, tumor, and purulent sinusitis. When worrisome symptoms and signs are present **(Table 16-2)**, rapid diagnosis and management are critical.

TABLE 16-2

Headache Symptoms That Suggest a Serious Underlying Disorder

Sudden-onset headache
First severe headache
“Worst” headache ever
Vomiting that precedes headache
Subacute worsening over days or weeks
Pain induced by bending, lifting, coughing
Pain that disturbs sleep or presents immediately upon awakening
Known systemic illness
Onset after age 55
Pain associated with local tenderness, e.g., region of temporal artery

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 16-2

Headache Symptoms That Suggest a Serious Underlying Disorder

Sudden-onset headache
First severe headache
“Worst” headache ever
Vomiting that precedes headache
Subacute worsening over days or weeks
Pain induced by bending, lifting, coughing
Pain that disturbs sleep or presents immediately upon awakening
Known systemic illness
Onset after age 55
Fever or unexplained systemic signs
Abnormal neurologic examination
Pain associated with local tenderness, e.g., region of temporal artery

A careful neurologic examination is an essential first step in the evaluation. In most cases, patients with an abnormal examination or a history of recent-onset headache should be evaluated by a computed tomography (CT) or magnetic resonance imaging (MRI) study of the brain. As an initial screening procedure for intracranial pathology in this setting, CT and MRI methods appear to be equally sensitive. In some circumstances, a lumbar puncture (LP) is also required, unless a benign etiology can be otherwise established. A general evaluation of acute headache might include cranial arteries by palpation; cervical spine by the effect of passive movement of the head and by imaging; the investigation of cardiovascular and renal status by blood pressure monitoring and urine examination; and eyes by funduscopy, intraocular pressure measurement, and refraction.

The patient’s psychological state should also be evaluated because a relationship exists between head pain, depression, and anxiety. This is intended to identify comorbidity rather than provide an explanation for the headache, because troublesome headache is seldom simply caused by mood change. Although it is notable that medicines with antidepressant actions are also effective in the preventive treatment of both tension-type headache and migraine, each symptom must be treated optimally.

Underlying recurrent headache disorders may be activated by pain that follows otologic or endodontic surgical procedures. Thus, pain about the head as the result of diseased tissue or trauma may reawaken an otherwise quiescent migraine syndrome. Treatment of the headache is largely ineffective until the cause of the primary problem is addressed.

Serious underlying conditions that are associated with headache are described below. Brain tumor is a rare cause of headache and even less commonly a cause of severe pain. The vast majority of patients presenting with severe headache have a benign cause.

SECONDARY HEADACHE

The management of secondary headache focuses on diagnosis and treatment of the underlying condition.

MENINGITIS

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

frequently present, perhaps reflecting the underlying biology of some of the patients.

MENINGITIS

Acute, severe headache with stiff neck and fever suggests meningitis. LP is mandatory. Often there is striking accentuation of pain with eye movement. Meningitis can be easily mistaken for migraine in that the cardinal symptoms of pounding headache, photophobia, nausea, and vomiting are frequently present, perhaps reflecting the underlying biology of some of the patients.

Meningitis is discussed in Chaps. 138 and 139.

INTRACRANIAL HEMORRHAGE

Acute, maximal in <5 min, severe headache lasting >5 min with stiff neck but without fever suggests subarachnoid hemorrhage. A ruptured aneurysm, arteriovenous malformation, or intraparenchymal hemorrhage may also present with headache alone. Rarely, if the hemorrhage is small or below the foramen magnum, the head CT scan can be normal. Therefore, LP may be required to diagnose definitively subarachnoid hemorrhage.

Subarachnoid hemorrhage is discussed in Chap. 429, and intracranial hemorrhage in Chap. 428.

BRAIN TUMOR

Approximately 30% of patients with brain tumors consider headache to be their chief complaint. The head pain is usually nondescript—an intermittent deep, dull aching of moderate intensity, which may worsen with exertion or change in position and may be associated with nausea and vomiting. This pattern of symptoms results from migraine far more often than from brain tumor. The headache of brain tumor disturbs sleep in about 10% of patients. Vomiting that precedes the appearance of headache by weeks is highly characteristic of posterior fossa brain tumors. A history of amenorrhea or galactorrhea should lead one to question whether a prolactin-secreting pituitary adenoma (or polycystic ovary syndrome) is the source of headache. Headache arising de novo in a patient with known malignancy suggests either cerebral metastases or carcinomatous meningitis. Head pain appearing abruptly after bending, lifting, or coughing can be due to a posterior fossa mass, a Chiari malformation, or low cerebrospinal fluid (CSF) volume.

Brain tumors are discussed in Chap. 90.

TEMPORAL ARTERITIS

Temporal (giant cell) arteritis is an inflammatory disorder of arteries that frequently involves the extracranial carotid circulation **(SEE ALSO Chaps. 32 AND 363)**. It is a common disorder of the elderly; its annual incidence is 77 per 100,000 individuals aged ≥50. The average age of onset is 70 years, and women account for 65% of cases. About half of patients with untreated temporal arteritis develop blindness due to involvement of the ophthalmic artery and its branches; indeed, the ischemic optic neuropathy induced by giant cell arteritis is the major cause of rapidly developing bilateral blindness in patients >60 years. Because treatment with glucocorticoids is effective in preventing this complication, prompt recognition of the disorder is important.

Typical presenting symptoms include headache, polymyalgia rheumatica **(Chap. 363)**, jaw claudication, fever, and weight loss. Headache is the dominant symptom and often appears in association with malaise and muscle aches. Head pain may be unilateral or bilateral and is located temporally in 50% of patients but may involve any and all aspects of the cranium. Pain usually appears gradually over a few hours before peak intensity is reached; occasionally, it is explosive in onset. The quality of pain is infrequently throbbing; it is almost invariably described as dull and boring, with superimposed episodic stabbing pains similar to the sharp pains that appear in migraine. Most patients can recognize that the origin of their head pain is superficial, external to the skull, rather than originating deep within the cranium (the pain site usually identified by migraineurs). Scalp tenderness is present, often to a marked degree; brushing the hair or resting the head on a pillow may be impossible because of pain. Headache is usually worse at night and often aggravated by exposure to cold. Additional findings may include reddened, tender nodules or red streaking of the skin overlying the temporal arteries, and tenderness of the temporal or, less commonly, the occipital arteries.

The erythrocyte sedimentation rate (ESR) is often, although not always, elevated; a normal ESR does not exclude giant cell arteritis. A temporal artery biopsy followed by immediate treatment with prednisone 80 mg daily for the first 4–6 weeks should be initiated when clinical suspicion is high; treatment should not be unreasonably delayed to obtain a biopsy. The prevalence of migraine among the elderly is substantial, considerably higher than that of giant cell arteritis. Migraineurs often report amelioration of their headache with prednisone; thus, caution must be used when interpreting the therapeutic response.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Glaucoma may present with a prostrating headache associated with nausea and vomiting. The headache often starts with severe eye pain. On physical examination, the eye is often red with a fixed, moderately dilated pupil.

ACCESS Medicine®
Access Provided by: SILVERCHAIR

than that of giant cell arteritis. Migraineurs often report amelioration of their headache with prednisone; thus, caution must be used when interpreting the therapeutic response.

GLAUCOMA

Glaucoma may present with a prostrating headache associated with nausea and vomiting. The headache often starts with severe eye pain. On physical examination, the eye is often red with a fixed, moderately dilated pupil.

Glaucoma is discussed in Chap. 32.

PRIMARY HEADACHE DISORDERS

Primary headaches are disorders in which headache and associated features occur in the absence of any exogenous cause. The most common are migraine, tension-type headache, and the TACs, notably cluster headache. These entities are discussed in detail in **Chap. 430**.

CHRONIC DAILY OR NEAR-DAILY HEADACHE

The broad description of chronic daily headache (CDH) can be applied when a patient experiences headache on 15 days or more per month. CDH is neither a single entity nor a diagnosis; it encompasses a number of different headache syndromes, both primary and secondary **(Table 16-3)**. In aggregate, this group presents considerable disability and is thus specially mentioned here. Population-based estimates suggest that about 4% of adults have daily or near-daily headache.

TABLE 16-3

Classification of Daily or Near-Daily Headache

Primary		
>4 H DAILY	**<4 H DAILY**	**SECONDARY**
Chronic migraine[a]	Chronic cluster headache[b]	Posttraumatic Head injury Iatrogenic Postinfectious
Chronic tension-type headache[a]	Chronic paroxysmal hemicrania	Inflammatory, such as Giant cell arteritis Sarcoidosis Behçet's syndrome
Hemicrania continua[a]	SUNCT/SUNA	Chronic CNS infection
New daily persistent headache[a]	Hypnic headache	Medication-overuse headache[a]

[a]May be complicated by medication overuse. [b]Some patients may have headache >4 h/d.

Abbreviations: CNS, central nervous system; SUNA, *s*hort-lasting *u*nilateral *n*euralgiform headache *a*ttacks with cranial autonomic symptoms; SUNCT, *s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing.

APPROACH TO THE PATIENT

Chronic Daily Headache

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

the headache type will guide therapy. Preventive treatments such as tricyclics, either amitriptyline or nortriptyline, at doses up to 1 mg/kg, are very useful in patients with CDH arising from migraine or tension-type headache or where the secondary cause has activated the underlying primary

unilateral neuralgiform headache attacks with conjunctival injection and tearing.

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

The first step in the management of patients with CDH is to diagnose any secondary headache and treat that problem **(Table 16-3)**. This can sometimes be a challenge when the underlying cause triggers worsening of a primary headache. For patients with primary headaches, diagnosis of the headache type will guide therapy. Preventive treatments such as tricyclics, either amitriptyline or nortriptyline, at doses up to 1 mg/kg, are very useful in patients with CDH arising from migraine or tension-type headache or where the secondary cause has activated the underlying primary headache. Tricyclics are started in low doses (10–25 mg daily) and may be given 12 h before the expected time of awakening in order to avoid excessive morning sleepiness. Medicines including topiramate, valproate, propranolol, flunarizine (not available in the United States), candesartan, and the newer calcitonin gene-related peptide (CGRP) pathway monoclonal antibodies, or gepants-CGRP receptor antagonists (see **Chap. 430**) are also useful when the underlying issue is migraine.

MANAGEMENT OF MEDICALLY INTRACTABLE DISABLING PRIMARY HEADACHE

The management of medically intractable headache is difficult, although recent developments in therapy are at hand. Monoclonal antibodies to CGRP or its receptor have been reported to be effective and well tolerated in chronic migraine and are now licensed for use in clinical practice. Noninvasive neuromodulatory approaches, such as single-pulse transcranial magnetic stimulation and noninvasive vagal nerve stimulation, which appear to modulate thalamic processing or brainstem mechanisms, respectively, in migraine have been used in clinical practice with success. Noninvasive vagal nerve stimulation has also shown promise particularly in chronic cluster headache, chronic paroxysmal hemicrania, and hemicrania continua, and possibly in short-lasting unilateral neuralgiform headache attacks with cranial autonomic symptoms (SUNA) and short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing (SUNCT) **(Chap. 430)**. Other modalities are discussed in **Chap. 430**.

MEDICATION-RELATED AND MEDICATION-OVERUSE HEADACHE

Overuse of analgesic medication for headache can aggravate headache frequency, markedly impair the effect of preventive medicines, and induce a state of refractory daily or near-daily headache called *medication-overuse headache*. A proportion of patients who stop taking analgesics will experience substantial improvement in the severity and frequency of their headache. However, even after cessation of analgesic use, many patients continue to have headache, although they may feel clinically improved in some way, especially if they have been using opioids or barbiturates regularly. The residual symptoms probably represent the underlying primary headache disorder, and most commonly this issue occurs in patients prone to migraine.

Management of Medication Overuse: Outpatients

For patients who overuse analgesic medications, it is often helpful to reduce and eliminate the medications, although this approach is far from universally effective. One approach is to reduce the medication dose by 10% every 1–2 weeks. Immediate cessation of analgesic use is possible for some patients, provided there is no contraindication. Both approaches are facilitated by use of a medication diary maintained during the month or two before cessation; this helps to identify the scope of the problem. A small dose of a nonsteroidal anti-inflammatory drug (NSAID) such as naproxen, 500 mg bid, if tolerated, will help relieve residual pain as analgesic use is reduced. NSAID overuse is not usually a problem for patients with daily headache when an NSAID with a longer half-life is taken once or twice daily; however, overuse problems may develop with shorter-acting NSAIDS. Once the patient has substantially reduced analgesic use, a preventive medication should be introduced. Another widely used approach is to commence the preventive at the same time the analgesic reduction is started. It must be emphasized that *preventives may not work in the presence of analgesic overuse, particularly with opioids*. The most common cause of unresponsiveness to treatment is the use of a preventive when analgesics continue to be used regularly. For some patients, discontinuing analgesics is very difficult; often the best approach is to inform the patient that some degree of headache is inevitable during this initial period.

Management of Medication Overuse: Inpatients

Some patients will require hospitalization for detoxification. Such patients have typically failed efforts at outpatient withdrawal or have a significant medical condition, such as diabetes mellitus or epilepsy, which would complicate withdrawal as an outpatient. Following admission to the hospital, medications are withdrawn completely on the first day, in the absence of a contraindication. Antiemetics and fluids are administered as required; clonidine is used for opioid withdrawal symptoms. For acute intolerable pain during the waking hours, aspirin, 1 g IV (not approved in the United States), is useful. IM chlorpromazine can be helpful at night; patients must be adequately hydrated. Three to five days into the admission, as the effect of the withdrawn substance wears off, a course of IV dihydroergotamine (DHE) can be used. DHE, administered every 8 h for 5 consecutive days, a treatment that is not stopped short if headache settles, can induce a significant remission that allows a preventive treatment to be established. Serotonin 5-HT_3 receptor antagonists, such as ondansetron or granisetron, or the neurokinin receptor antagonist, aprepitant, may be required with DHE to prevent significant nausea, and domperidone (not approved in the United States) orally or by suppository can be very helpful

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

New daily persistent headache (NDPH) is a clinically distinct syndrome with important secondary causes; these are listed in **Table 16-4**.

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

clonidine is used for opioid withdrawal symptoms. For acute intolerable pain during the waking hours, aspirin, 1 g IV (not approved in the United

established. Serotonin 5-HT_3 receptor antagonists, such as ondansetron or granisetron, or the neurokinin receptor antagonist, aprepitant, may be required with DHE to prevent significant nausea, and domperidone (not approved in the United States) orally or by suppository can be very helpful. Avoiding sedating or otherwise side effect–prone antiemetics is helpful.

NEW DAILY PERSISTENT HEADACHE

New daily persistent headache (NDPH) is a clinically distinct syndrome with important secondary causes; these are listed in **Table 16-4**.

Clinical Presentation

NDPH presents with headache on most if not all days, and the patient can clearly, and often vividly, recall the moment of onset. The headache usually begins abruptly, but onset may be more gradual; evolution over 3 days has been proposed as the upper limit for this syndrome. Patients typically recall the exact day and circumstances of the onset of headache; the new, persistent head pain does not remit. The first priority is to distinguish between a primary and a secondary cause of this syndrome. Subarachnoid hemorrhage is the most serious of the secondary causes and must be excluded either by history or appropriate investigation **(Chap. 429)**.

Secondary NDPH

Low CSF Volume Headache In these syndromes, head pain is positional: it begins when the patient sits or stands upright and resolves upon reclining. The pain, which is occipitofrontal, is usually a dull ache but may be throbbing. Patients with chronic low CSF volume headache typically present with a history of headache from one day to the next that is generally not present on waking but worsens during the day. Recumbency usually improves the headache within minutes, and it can take only minutes to an hour for the pain to return when the patient resumes an upright position.

The most common cause of headache due to persistent low CSF volume is CSF leak following LP **(Chap. S9)**. Post-LP headache usually begins within 48 h but may be delayed for up to 12 days. Its incidence is between 10% and 30%. Beverages with caffeine may provide temporary relief. Besides LP, index events may include epidural injection or a vigorous Valsalva maneuver, such as from lifting, straining, coughing, clearing the eustachian tubes in an airplane, or multiple orgasms. Spontaneous CSF leaks are well recognized, and the diagnosis should be considered whenever the headache history is typical, even when there is no obvious index event. As time passes from the index event, the postural nature may become less apparent; cases in which the index event occurred several years before the eventual diagnosis have been recognized. Symptoms appear to result from low volume rather than low pressure: although low CSF pressures, typically 0–50 mm CSF, are usually identified, a pressure as high as 140 mm CSF has been noted with a documented leak.

Postural orthostatic tachycardia syndrome (POTS; **Chap. 440**) can present with orthostatic headache similar to low CSF volume headache and is a diagnosis that needs consideration in this setting.

When imaging is indicated to identify the source of a presumed leak, an MRI with gadolinium is the initial study of choice **(Fig. 16-1)**. A striking pattern of diffuse meningeal enhancement is so typical that in the appropriate clinical context the diagnosis is established. Chiari malformations may sometimes be noted on MRI; in such cases, surgery to decompress the posterior fossa is *not* indicated and usually worsens the headache. Spinal MRI with T2 weighting may reveal a leak, and spinal MRI may demonstrate spinal meningeal cysts whose role in these syndromes is yet to be elucidated. The source of CSF leakage may be identified by spinal MRI with appropriate sequences, or by CT, preferably digital subtraction, myelography. In the absence of a directly identified site of leakage, 111In-DTPA CSF studies may demonstrate early emptying of the tracer into the bladder or slow progress of tracer across the brain suggesting a CSF leak; this procedure is now only rarely employed.

Initial treatment for low CSF volume headache is bed rest. For patients with persistent pain, IV caffeine (500 mg in 500 mL of saline administered over 2 h) can be very effective. An electrocardiogram (ECG) to screen for arrhythmia should be performed before administration. It is reasonable to administer at least two infusions of caffeine before embarking on additional tests to identify the source of the CSF leak. Because IV caffeine is safe and can be curative, it spares many patients the need for further investigations. If unsuccessful, an abdominal binder may be helpful. If a leak can be identified, an autologous blood patch is usually curative. A blood patch is also effective for post-LP headache; in this setting, the location is empirically determined to be the site of the LP. In patients with intractable headache, oral theophylline is a useful alternative that can take some months to be effective.

Raised CSF Pressure Headache

Raised CSF pressure is well recognized as a cause of headache. Brain imaging can often reveal the cause, such as a space-occupying lesion.

Idiopathic intracranial hypertension (pseudotumor cerebri) NDPH due to raised CSF pressure can be the presenting symptom for patients with idiopathic intracranial hypertension, a disorder associated with obesity, female gender, and, on occasion, pregnancy. The syndrome can also

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

is present, but the possibility must be considered even in patients without funduscopic changes. Formal visual field testing should be performed even in the absence of overt ophthalmic involvement. Partial obstructions of the cerebral venous sinuses are found in a small number of cases. In

with idiopathic intracranial hypertension, a disorder associated with obesity, female gender, and, on occasion, pregnancy. The syndrome can also occur without visual problems, particularly when the fundi are normal. These patients typically present with a history of generalized headache that is present on waking and improves as the day goes on. It is generally present on awakening in the morning and is worse with recumbency. Transient visual obscurations are frequent and may occur when the headaches are most severe. The diagnosis is relatively straightforward when papilledema is present, but the possibility must be considered even in patients without funduscopic changes. Formal visual field testing should be performed even in the absence of overt ophthalmic involvement. Partial obstructions of the cerebral venous sinuses are found in a small number of cases. In addition, persistently raised intracranial pressure can trigger a syndrome of chronic migraine. Other conditions that characteristically produce headache on rising in the morning or nocturnal headache are obstructive sleep apnea or poorly controlled hypertension.

Evaluation of patients suspected to have raised CSF pressure requires brain imaging. It is most efficient to obtain an MRI, including an MR venogram, as the initial study. If there are no contraindications, the CSF pressure should be measured by LP; this should be done when the patient is symptomatic so that both the pressure and the response to removal of 20–30 mL of CSF can be determined. An elevated opening pressure and improvement in headache following removal of CSF are diagnostic in the absence of fundal changes.

Initial treatment is with acetazolamide (250–500 mg bid); the headache may improve within weeks. If ineffective, topiramate is the next treatment of choice; it has many actions that may be useful in this setting, including carbonic anhydrase inhibition, weight loss, and neuronal membrane stabilization, likely mediated via effects on phosphorylation pathways. Severely disabled patients who do not respond to medical treatment require intracranial pressure monitoring and may require shunting. If appropriate, weight loss should be encouraged.

Posttraumatic Headache

A traumatic event can trigger a headache process that lasts for many months or years after the event. The term *trauma* is used here in a very broad sense: headache can develop following an injury to the head, but it can also develop after an infectious episode, typically viral meningitis; a flulike illness; or a parasitic infection. Complaints of dizziness, vertigo, and impaired memory can accompany the headache. Symptoms may remit after several weeks or persist for months and even years after the injury. Typically, the neurologic examination is normal and CT or MRI studies are unrevealing. Chronic subdural hematoma may on occasion mimic this disorder. Posttraumatic headache may also be seen after carotid dissection and subarachnoid hemorrhage and after intracranial surgery. The underlying theme appears to be that a traumatic event involving the pain-producing meninges can trigger a headache process that lasts for many years.

Other Causes

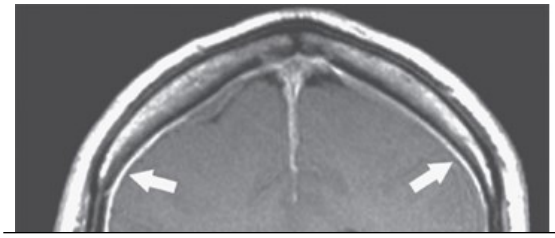
In one series, one-third of patients with NDPH reported headache beginning after a transient flulike illness characterized by fever, neck stiffness, photophobia, and marked malaise. Evaluation typically reveals no apparent cause for the headache. There is no convincing evidence that persistent Epstein-Barr virus infection plays a role in NDPH. A complicating factor is that many patients undergo LP during the acute illness; iatrogenic low CSF volume headache must be considered in these cases.

Treatment

Treatment is largely empirical and directed at the headache phenotype. Tricyclic antidepressants, notably amitriptyline, and anticonvulsants, such as topiramate, valproate, candesartan, and gabapentin, have been used with reported benefit. The monoamine oxidase inhibitor phenelzine may also be useful in carefully selected patients. The headache usually resolves within 3–5 years, but it can be quite disabling.

FIGURE 16-1

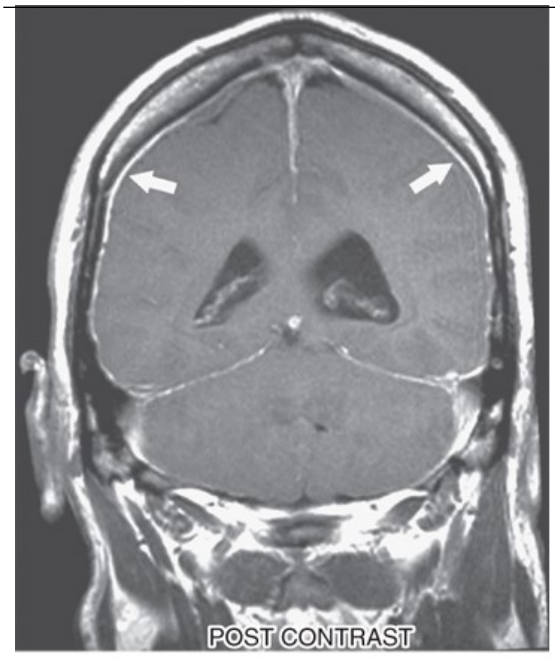
Magnetic resonance image showing diffuse meningeal enhancement after gadolinium administration in a patient with low cerebrospinal fluid (CSF) volume headache. (*Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.*)



Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Edition Copyright © McGraw Hill Education. All rights reserved.)

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 16-4

Differential Diagnosis of New Daily Persistent Headache

PRIMARY	SECONDARY
Migrainous-type	Subarachnoid hemorrhage
Featureless (tension-type)	Low cerebrospinal fluid (CSF) volume headache
	Raised CSF pressure headache
	Posttraumatic headache[a]
	Chronic meningitis

[a]Includes postinfectious forms.

PRIMARY CARE AND HEADACHE MANAGEMENT

Most patients with headache will be seen first in a primary care setting. The challenging task of the primary care physician is to identify the very few worrisome secondary headaches from the very great majority of primary and less dangerous secondary headaches (Table 16-2).

Absent any warning signs, a reasonable approach is to treat when a diagnosis is established. As a general rule, the investigation should focus on identifying worrisome causes of headache or on helping the patient to gain confidence if no primary headache diagnosis can be made.

After treatment has been initiated, follow-up care is essential to identify whether progress has been made against the headache complaint. Not all headaches will respond to treatment, but, in general, worrisome headaches will progress and will be easier to identify.

When a primary care physician feels the diagnosis is a primary headache disorder, it is worth noting that >90% of patients who present to primary care unresponsive to two or more standard therapies for the considered headache type, should be considered for referral to a specialist. In a practical sense, the threshold for referral is also determined by the experience of the primary care physician in headache medicine and the availability of

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Absent any warning signs, a reasonable approach is to treat when a diagnosis is established. As a general rule, the investigation should focus on headaches will respond to treatment, but, in general, worrisome headaches will progress and will be easier to identify.

When a primary care physician feels the diagnosis is a primary headache disorder, it is worth noting that >90% of patients who present to primary care with a complaint of headache will have migraine **(Chap. 430)**.

In general, patients who do not have a clear diagnosis, have a primary headache disorder other than migraine or tension-type headache, or are unresponsive to two or more standard therapies for the considered headache type, should be considered for referral to a specialist. In a practical sense, the threshold for referral is also determined by the experience of the primary care physician in headache medicine and the availability of secondary care options.

ACKNOWLEDGMENT

The editors acknowledge the contributions of Neil H. Raskin to earlier editions of this chapter.

FURTHER READING

Headache Classification Committee of the International Headache Society: The International Classification of Headache Disorders, 3rd ed. Cephalalgia 33:629, 2018.

Kernick D, Goadsby PJ: *Headache: A Practical Manual*. Oxford: Oxford University Press, 2008.

Lance JW, Goadsby PJ: *Mechanism and Management of Headache*, 7th ed. New York, Elsevier, 2005.

Olesen J et al: *The Headaches*. Philadelphia, Lippincott, Williams & Wilkins, 2005.

Silberstein SD, Lipton RB, Dodick DW: *Wolff's Headache and Other Head Pain*, 9th ed. New York, Oxford University Press, 2021.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

Chapter 16: Headache

Peter J. Goadsby

INTRODUCTION

Headache is among the most common reasons patients seek medical attention and is responsible, on a global basis, for more disability than any other neurologic problem. Diagnosis and management are based on a careful clinical approach augmented by an understanding of the anatomy, physiology, and pharmacology of the nervous system pathways mediating the various headache syndromes. This chapter will focus on the general approach to a patient with headache; migraine and other primary headache disorders are discussed in **Chap. 430**.

GENERAL PRINCIPLES

A classification system developed by the International Headache Society (*www.ihs-headache.org/en/resources/guidelines/*) characterizes headache as primary or secondary **(Table 16-1)**. *Primary headaches* are those in which headache and its associated features are the disorder itself, whereas *secondary headaches* are those caused by exogenous disorders (Headache Classification Committee of the International Headache Society, 2018). Primary headache often results in considerable disability and a decrease in the patient's quality of life. Mild secondary headache, such as that seen in association with upper respiratory tract infections, is common but rarely worrisome. Life-threatening headache is relatively uncommon, but vigilance is required in order to recognize and appropriately treat such patients.

TABLE 16-1

Common Causes of Headache

PRIMARY HEADACHE		SECONDARY HEADACHE	
TYPE	**%**	**TYPE**	**%**
Tension-type	69	Systemic infection	63
Migraine	16	Head injury	4
Idiopathic stabbing	2	Vascular disorders	1
Exertional	1	Subarachnoid hemorrhage	<1
Cluster	0.1	Brain tumor	0.1

Source: After J Olesen et al: *The Headaches*. Philadelphia, Lippincott Williams & Wilkins, 2005.

ANATOMY AND PHYSIOLOGY OF HEADACHE

Pain usually occurs when peripheral nociceptors are stimulated in response to tissue injury, visceral distension, or other factors **(Chap. 13)**. In such situations, pain perception is a normal physiologic response mediated by a healthy nervous system. Pain can also result when pain-producing pathways of the peripheral or central nervous system (CNS) are damaged or activated inappropriately. Headache may originate from either or both mechanisms. Relatively few cranial structures are pain producing; these include the scalp, meningeal arteries, dural sinuses, falx cerebri, and proximal

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

- The large intracranial vessels and dura mater, and the peripheral terminals of the trigeminal nerve that innervate these structures

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

pathways of the peripheral or central nervous system (CNS) are damaged or activated inappropriately. Headache may originate from either or both mechanisms. Relatively few cranial structures are pain producing; these include the scalp, meningeal arteries, dural sinuses, falx cerebri, and proximal segments of the large pial arteries. The ventricular ependyma, choroid plexus, pial veins, and much of the brain parenchyma are not pain producing.

The key structures involved in primary headache are the following:

- The large intracranial vessels and dura mater, and the peripheral terminals of the trigeminal nerve that innervate these structures
- The caudal portion of the trigeminal nucleus, which extends into the dorsal horns of the upper cervical spinal cord and receives input from the first and second cervical nerve roots (the trigeminocervical complex)
- Rostral pain-processing regions, such as the ventroposteromedial thalamus and the cortex
- The pain-modulatory systems in the brain that modulate input from the trigeminal nociceptors at all levels of the pain-processing pathways and influence vegetative functions, such as the hypothalamus and brainstem

The *trigeminovascular system* innervates the large intracranial vessels and dura mater via the trigeminal nerve. Cranial autonomic symptoms, such as lacrimation, conjunctival injection, nasal congestion, rhinorrhea, periorbital swelling, aural fullness, and ptosis, are prominent in the trigeminal autonomic cephalalgias (TACs), including cluster headache and paroxysmal hemicrania, and may also be seen in migraine, even in children. These autonomic symptoms reflect activation of cranial parasympathetic pathways, and functional imaging studies indicate that vascular changes in migraine and cluster headache, when present, are similarly driven by these cranial autonomic systems. Thus, they are secondary, and not causative, events in the headache cascade. Moreover, they can often be mistaken for symptoms or signs of cranial sinus inflammation, which is then overdiagnosed and inappropriately managed. Migraine and other primary headache types are not “vascular headaches”; these disorders do not reliably manifest vascular changes, and treatment outcomes cannot be predicted by vascular effects. Migraine is a brain disorder and is best understood and managed as such.

CLINICAL EVALUATION OF ACUTE, NEW-ONSET HEADACHE

The patient who presents with a new, severe headache has a differential diagnosis that is quite different from the patient with recurrent headaches over many years. In new-onset and severe headache, the probability of finding a potentially serious cause is considerably greater than in recurrent headache. Patients with recent onset of pain require prompt evaluation and appropriate treatment. Serious causes to be considered include meningitis, subarachnoid hemorrhage, epidural or subdural hematoma, glaucoma, tumor, and purulent sinusitis. When worrisome symptoms and signs are present **(Table 16-2)**, rapid diagnosis and management are critical.

TABLE 16-2

Headache Symptoms That Suggest a Serious Underlying Disorder

Sudden-onset headache
First severe headache
“Worst” headache ever
Vomiting that precedes headache
Subacute worsening over days or weeks
Pain induced by bending, lifting, coughing
Pain that disturbs sleep or presents immediately upon awakening
Known systemic illness
Onset after age 55
Pain associated with local tenderness, e.g., region of temporal artery

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 16-2

Headache Symptoms That Suggest a Serious Underlying Disorder

Sudden-onset headache
First severe headache
“Worst” headache ever
Vomiting that precedes headache
Subacute worsening over days or weeks
Pain induced by bending, lifting, coughing
Pain that disturbs sleep or presents immediately upon awakening
Known systemic illness
Onset after age 55
Fever or unexplained systemic signs
Abnormal neurologic examination
Pain associated with local tenderness, e.g., region of temporal artery

A careful neurologic examination is an essential first step in the evaluation. In most cases, patients with an abnormal examination or a history of recent-onset headache should be evaluated by a computed tomography (CT) or magnetic resonance imaging (MRI) study of the brain. As an initial screening procedure for intracranial pathology in this setting, CT and MRI methods appear to be equally sensitive. In some circumstances, a lumbar puncture (LP) is also required, unless a benign etiology can be otherwise established. A general evaluation of acute headache might include cranial arteries by palpation; cervical spine by the effect of passive movement of the head and by imaging; the investigation of cardiovascular and renal status by blood pressure monitoring and urine examination; and eyes by funduscopy, intraocular pressure measurement, and refraction.

The patient’s psychological state should also be evaluated because a relationship exists between head pain, depression, and anxiety. This is intended to identify comorbidity rather than provide an explanation for the headache, because troublesome headache is seldom simply caused by mood change. Although it is notable that medicines with antidepressant actions are also effective in the preventive treatment of both tension-type headache and migraine, each symptom must be treated optimally.

Underlying recurrent headache disorders may be activated by pain that follows otologic or endodontic surgical procedures. Thus, pain about the head as the result of diseased tissue or trauma may reawaken an otherwise quiescent migraine syndrome. Treatment of the headache is largely ineffective until the cause of the primary problem is addressed.

Serious underlying conditions that are associated with headache are described below. Brain tumor is a rare cause of headache and even less commonly a cause of severe pain. The vast majority of patients presenting with severe headache have a benign cause.

SECONDARY HEADACHE

The management of secondary headache focuses on diagnosis and treatment of the underlying condition.

MENINGITIS

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

frequently present, perhaps reflecting the underlying biology of some of the patients.

MENINGITIS

Acute, severe headache with stiff neck and fever suggests meningitis. LP is mandatory. Often there is striking accentuation of pain with eye movement. Meningitis can be easily mistaken for migraine in that the cardinal symptoms of pounding headache, photophobia, nausea, and vomiting are frequently present, perhaps reflecting the underlying biology of some of the patients.

Meningitis is discussed in Chaps. 138 and 139.

INTRACRANIAL HEMORRHAGE

Acute, maximal in <5 min, severe headache lasting >5 min with stiff neck but without fever suggests subarachnoid hemorrhage. A ruptured aneurysm, arteriovenous malformation, or intraparenchymal hemorrhage may also present with headache alone. Rarely, if the hemorrhage is small or below the foramen magnum, the head CT scan can be normal. Therefore, LP may be required to diagnose definitively subarachnoid hemorrhage.

Subarachnoid hemorrhage is discussed in Chap. 429, and intracranial hemorrhage in Chap. 428.

BRAIN TUMOR

Approximately 30% of patients with brain tumors consider headache to be their chief complaint. The head pain is usually nondescript—an intermittent deep, dull aching of moderate intensity, which may worsen with exertion or change in position and may be associated with nausea and vomiting. This pattern of symptoms results from migraine far more often than from brain tumor. The headache of brain tumor disturbs sleep in about 10% of patients. Vomiting that precedes the appearance of headache by weeks is highly characteristic of posterior fossa brain tumors. A history of amenorrhea or galactorrhea should lead one to question whether a prolactin-secreting pituitary adenoma (or polycystic ovary syndrome) is the source of headache. Headache arising de novo in a patient with known malignancy suggests either cerebral metastases or carcinomatous meningitis. Head pain appearing abruptly after bending, lifting, or coughing can be due to a posterior fossa mass, a Chiari malformation, or low cerebrospinal fluid (CSF) volume.

Brain tumors are discussed in Chap. 90.

TEMPORAL ARTERITIS

Temporal (giant cell) arteritis is an inflammatory disorder of arteries that frequently involves the extracranial carotid circulation **(SEE ALSO Chaps. 32 AND 363)**. It is a common disorder of the elderly; its annual incidence is 77 per 100,000 individuals aged ≥50. The average age of onset is 70 years, and women account for 65% of cases. About half of patients with untreated temporal arteritis develop blindness due to involvement of the ophthalmic artery and its branches; indeed, the ischemic optic neuropathy induced by giant cell arteritis is the major cause of rapidly developing bilateral blindness in patients >60 years. Because treatment with glucocorticoids is effective in preventing this complication, prompt recognition of the disorder is important.

Typical presenting symptoms include headache, polymyalgia rheumatica **(Chap. 363)**, jaw claudication, fever, and weight loss. Headache is the dominant symptom and often appears in association with malaise and muscle aches. Head pain may be unilateral or bilateral and is located temporally in 50% of patients but may involve any and all aspects of the cranium. Pain usually appears gradually over a few hours before peak intensity is reached; occasionally, it is explosive in onset. The quality of pain is infrequently throbbing; it is almost invariably described as dull and boring, with superimposed episodic stabbing pains similar to the sharp pains that appear in migraine. Most patients can recognize that the origin of their head pain is superficial, external to the skull, rather than originating deep within the cranium (the pain site usually identified by migraineurs). Scalp tenderness is present, often to a marked degree; brushing the hair or resting the head on a pillow may be impossible because of pain. Headache is usually worse at night and often aggravated by exposure to cold. Additional findings may include reddened, tender nodules or red streaking of the skin overlying the temporal arteries, and tenderness of the temporal or, less commonly, the occipital arteries.

The erythrocyte sedimentation rate (ESR) is often, although not always, elevated; a normal ESR does not exclude giant cell arteritis. A temporal artery biopsy followed by immediate treatment with prednisone 80 mg daily for the first 4–6 weeks should be initiated when clinical suspicion is high; treatment should not be unreasonably delayed to obtain a biopsy. The prevalence of migraine among the elderly is substantial, considerably higher than that of giant cell arteritis. Migraineurs often report amelioration of their headache with prednisone; thus, caution must be used when interpreting the therapeutic response.

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Glaucoma may present with a prostrating headache associated with nausea and vomiting. The headache often starts with severe eye pain. On physical examination, the eye is often red with a fixed, moderately dilated pupil.

ACCESS Medicine®
Access Provided by: SILVERCHAIR

than that of giant cell arteritis. Migraineurs often report amelioration of their headache with prednisone; thus, caution must be used when interpreting the therapeutic response.

GLAUCOMA

Glaucoma may present with a prostrating headache associated with nausea and vomiting. The headache often starts with severe eye pain. On physical examination, the eye is often red with a fixed, moderately dilated pupil.

Glaucoma is discussed in Chap. 32.

PRIMARY HEADACHE DISORDERS

Primary headaches are disorders in which headache and associated features occur in the absence of any exogenous cause. The most common are migraine, tension-type headache, and the TACs, notably cluster headache. These entities are discussed in detail in **Chap. 430**.

CHRONIC DAILY OR NEAR-DAILY HEADACHE

The broad description of chronic daily headache (CDH) can be applied when a patient experiences headache on 15 days or more per month. CDH is neither a single entity nor a diagnosis; it encompasses a number of different headache syndromes, both primary and secondary **(Table 16-3)**. In aggregate, this group presents considerable disability and is thus specially mentioned here. Population-based estimates suggest that about 4% of adults have daily or near-daily headache.

TABLE 16-3

Classification of Daily or Near-Daily Headache

Primary		
>4 H DAILY	**<4 H DAILY**	**SECONDARY**
Chronic migraine[a]	Chronic cluster headache[b]	Posttraumatic Head injury Iatrogenic Postinfectious
Chronic tension-type headache[a]	Chronic paroxysmal hemicrania	Inflammatory, such as Giant cell arteritis Sarcoidosis Behçet's syndrome
Hemicrania continua[a]	SUNCT/SUNA	Chronic CNS infection
New daily persistent headache[a]	Hypnic headache	Medication-overuse headache[a]

[a]May be complicated by medication overuse. [b]Some patients may have headache >4 h/d.

Abbreviations: CNS, central nervous system; SUNA, *s*hort-lasting *u*nilateral *n*euralgiform headache *a*ttacks with cranial autonomic symptoms; SUNCT, *s*hort-lasting *u*nilateral *n*euralgiform headache attacks with *c*onjunctival injection and *t*earing.

APPROACH TO THE PATIENT

Chronic Daily Headache

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

the headache type will guide therapy. Preventive treatments such as tricyclics, either amitriptyline or nortriptyline, at doses up to 1 mg/kg, are very useful in patients with CDH arising from migraine or tension-type headache or where the secondary cause has activated the underlying primary

unilateral neuralgiform headache attacks with conjunctival injection and tearing.

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

The first step in the management of patients with CDH is to diagnose any secondary headache and treat that problem **(Table 16-3)**. This can sometimes be a challenge when the underlying cause triggers worsening of a primary headache. For patients with primary headaches, diagnosis of the headache type will guide therapy. Preventive treatments such as tricyclics, either amitriptyline or nortriptyline, at doses up to 1 mg/kg, are very useful in patients with CDH arising from migraine or tension-type headache or where the secondary cause has activated the underlying primary headache. Tricyclics are started in low doses (10–25 mg daily) and may be given 12 h before the expected time of awakening in order to avoid excessive morning sleepiness. Medicines including topiramate, valproate, propranolol, flunarizine (not available in the United States), candesartan, and the newer calcitonin gene-related peptide (CGRP) pathway monoclonal antibodies, or gepants-CGRP receptor antagonists (see **Chap. 430**) are also useful when the underlying issue is migraine.

MANAGEMENT OF MEDICALLY INTRACTABLE DISABLING PRIMARY HEADACHE

The management of medically intractable headache is difficult, although recent developments in therapy are at hand. Monoclonal antibodies to CGRP or its receptor have been reported to be effective and well tolerated in chronic migraine and are now licensed for use in clinical practice. Noninvasive neuromodulatory approaches, such as single-pulse transcranial magnetic stimulation and noninvasive vagal nerve stimulation, which appear to modulate thalamic processing or brainstem mechanisms, respectively, in migraine have been used in clinical practice with success. Noninvasive vagal nerve stimulation has also shown promise particularly in chronic cluster headache, chronic paroxysmal hemicrania, and hemicrania continua, and possibly in short-lasting unilateral neuralgiform headache attacks with cranial autonomic symptoms (SUNA) and short-lasting unilateral neuralgiform headache attacks with conjunctival injection and tearing (SUNCT) **(Chap. 430)**. Other modalities are discussed in **Chap. 430**.

MEDICATION-RELATED AND MEDICATION-OVERUSE HEADACHE

Overuse of analgesic medication for headache can aggravate headache frequency, markedly impair the effect of preventive medicines, and induce a state of refractory daily or near-daily headache called *medication-overuse headache*. A proportion of patients who stop taking analgesics will experience substantial improvement in the severity and frequency of their headache. However, even after cessation of analgesic use, many patients continue to have headache, although they may feel clinically improved in some way, especially if they have been using opioids or barbiturates regularly. The residual symptoms probably represent the underlying primary headache disorder, and most commonly this issue occurs in patients prone to migraine.

Management of Medication Overuse: Outpatients

For patients who overuse analgesic medications, it is often helpful to reduce and eliminate the medications, although this approach is far from universally effective. One approach is to reduce the medication dose by 10% every 1–2 weeks. Immediate cessation of analgesic use is possible for some patients, provided there is no contraindication. Both approaches are facilitated by use of a medication diary maintained during the month or two before cessation; this helps to identify the scope of the problem. A small dose of a nonsteroidal anti-inflammatory drug (NSAID) such as naproxen, 500 mg bid, if tolerated, will help relieve residual pain as analgesic use is reduced. NSAID overuse is not usually a problem for patients with daily headache when an NSAID with a longer half-life is taken once or twice daily; however, overuse problems may develop with shorter-acting NSAIDS. Once the patient has substantially reduced analgesic use, a preventive medication should be introduced. Another widely used approach is to commence the preventive at the same time the analgesic reduction is started. It must be emphasized that *preventives may not work in the presence of analgesic overuse, particularly with opioids*. The most common cause of unresponsiveness to treatment is the use of a preventive when analgesics continue to be used regularly. For some patients, discontinuing analgesics is very difficult; often the best approach is to inform the patient that some degree of headache is inevitable during this initial period.

Management of Medication Overuse: Inpatients

Some patients will require hospitalization for detoxification. Such patients have typically failed efforts at outpatient withdrawal or have a significant medical condition, such as diabetes mellitus or epilepsy, which would complicate withdrawal as an outpatient. Following admission to the hospital, medications are withdrawn completely on the first day, in the absence of a contraindication. Antiemetics and fluids are administered as required; clonidine is used for opioid withdrawal symptoms. For acute intolerable pain during the waking hours, aspirin, 1 g IV (not approved in the United States), is useful. IM chlorpromazine can be helpful at night; patients must be adequately hydrated. Three to five days into the admission, as the effect of the withdrawn substance wears off, a course of IV dihydroergotamine (DHE) can be used. DHE, administered every 8 h for 5 consecutive days, a treatment that is not stopped short if headache settles, can induce a significant remission that allows a preventive treatment to be established. Serotonin 5-HT_3 receptor antagonists, such as ondansetron or granisetron, or the neurokinin receptor antagonist, aprepitant, may be required with DHE to prevent significant nausea, and domperidone (not approved in the United States) orally or by suppository can be very helpful.

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

New daily persistent headache (NDPH) is a clinically distinct syndrome with important secondary causes; these are listed in **Table 16-4**.

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

clonidine is used for opioid withdrawal symptoms. For acute intolerable pain during the waking hours, aspirin, 1 g IV (not approved in the United

established. Serotonin 5-HT_3 receptor antagonists, such as ondansetron or granisetron, or the neurokinin receptor antagonist, aprepitant, may be required with DHE to prevent significant nausea, and domperidone (not approved in the United States) orally or by suppository can be very helpful. Avoiding sedating or otherwise side effect–prone antiemetics is helpful.

NEW DAILY PERSISTENT HEADACHE

New daily persistent headache (NDPH) is a clinically distinct syndrome with important secondary causes; these are listed in **Table 16-4**.

Clinical Presentation

NDPH presents with headache on most if not all days, and the patient can clearly, and often vividly, recall the moment of onset. The headache usually begins abruptly, but onset may be more gradual; evolution over 3 days has been proposed as the upper limit for this syndrome. Patients typically recall the exact day and circumstances of the onset of headache; the new, persistent head pain does not remit. The first priority is to distinguish between a primary and a secondary cause of this syndrome. Subarachnoid hemorrhage is the most serious of the secondary causes and must be excluded either by history or appropriate investigation **(Chap. 429)**.

Secondary NDPH

Low CSF Volume Headache In these syndromes, head pain is positional: it begins when the patient sits or stands upright and resolves upon reclining. The pain, which is occipitofrontal, is usually a dull ache but may be throbbing. Patients with chronic low CSF volume headache typically present with a history of headache from one day to the next that is generally not present on waking but worsens during the day. Recumbency usually improves the headache within minutes, and it can take only minutes to an hour for the pain to return when the patient resumes an upright position.

The most common cause of headache due to persistent low CSF volume is CSF leak following LP **(Chap. S9)**. Post-LP headache usually begins within 48 h but may be delayed for up to 12 days. Its incidence is between 10% and 30%. Beverages with caffeine may provide temporary relief. Besides LP, index events may include epidural injection or a vigorous Valsalva maneuver, such as from lifting, straining, coughing, clearing the eustachian tubes in an airplane, or multiple orgasms. Spontaneous CSF leaks are well recognized, and the diagnosis should be considered whenever the headache history is typical, even when there is no obvious index event. As time passes from the index event, the postural nature may become less apparent; cases in which the index event occurred several years before the eventual diagnosis have been recognized. Symptoms appear to result from low volume rather than low pressure: although low CSF pressures, typically 0–50 mm CSF, are usually identified, a pressure as high as 140 mm CSF has been noted with a documented leak.

Postural orthostatic tachycardia syndrome (POTS; **Chap. 440**) can present with orthostatic headache similar to low CSF volume headache and is a diagnosis that needs consideration in this setting.

When imaging is indicated to identify the source of a presumed leak, an MRI with gadolinium is the initial study of choice **(Fig. 16-1)**. A striking pattern of diffuse meningeal enhancement is so typical that in the appropriate clinical context the diagnosis is established. Chiari malformations may sometimes be noted on MRI; in such cases, surgery to decompress the posterior fossa is *not* indicated and usually worsens the headache. Spinal MRI with T2 weighting may reveal a leak, and spinal MRI may demonstrate spinal meningeal cysts whose role in these syndromes is yet to be elucidated. The source of CSF leakage may be identified by spinal MRI with appropriate sequences, or by CT, preferably digital subtraction, myelography. In the absence of a directly identified site of leakage, 111In-DTPA CSF studies may demonstrate early emptying of the tracer into the bladder or slow progress of tracer across the brain suggesting a CSF leak; this procedure is now only rarely employed.

Initial treatment for low CSF volume headache is bed rest. For patients with persistent pain, IV caffeine (500 mg in 500 mL of saline administered over 2 h) can be very effective. An electrocardiogram (ECG) to screen for arrhythmia should be performed before administration. It is reasonable to administer at least two infusions of caffeine before embarking on additional tests to identify the source of the CSF leak. Because IV caffeine is safe and can be curative, it spares many patients the need for further investigations. If unsuccessful, an abdominal binder may be helpful. If a leak can be identified, an autologous blood patch is usually curative. A blood patch is also effective for post-LP headache; in this setting, the location is empirically determined to be the site of the LP. In patients with intractable headache, oral theophylline is a useful alternative that can take some months to be effective.

Raised CSF Pressure Headache

Raised CSF pressure is well recognized as a cause of headache. Brain imaging can often reveal the cause, such as a space-occupying lesion.

Idiopathic intracranial hypertension (pseudotumor cerebri) NDPH due to raised CSF pressure can be the presenting symptom for patients with idiopathic intracranial hypertension, a disorder associated with obesity, female gender, and, on occasion, pregnancy. The syndrome can also

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

is present, but the possibility must be considered even in patients without funduscopic changes. Formal visual field testing should be performed even in the absence of overt ophthalmic involvement. Partial obstructions of the cerebral venous sinuses are found in a small number of cases. In

with idiopathic intracranial hypertension, a disorder associated with obesity, female gender, and, on occasion, pregnancy. The syndrome can also occur without visual problems, particularly when the fundi are normal. These patients typically present with a history of generalized headache that is present on waking and improves as the day goes on. It is generally present on awakening in the morning and is worse with recumbency. Transient visual obscurations are frequent and may occur when the headaches are most severe. The diagnosis is relatively straightforward when papilledema is present, but the possibility must be considered even in patients without funduscopic changes. Formal visual field testing should be performed even in the absence of overt ophthalmic involvement. Partial obstructions of the cerebral venous sinuses are found in a small number of cases. In addition, persistently raised intracranial pressure can trigger a syndrome of chronic migraine. Other conditions that characteristically produce headache on rising in the morning or nocturnal headache are obstructive sleep apnea or poorly controlled hypertension.

Evaluation of patients suspected to have raised CSF pressure requires brain imaging. It is most efficient to obtain an MRI, including an MR venogram, as the initial study. If there are no contraindications, the CSF pressure should be measured by LP; this should be done when the patient is symptomatic so that both the pressure and the response to removal of 20–30 mL of CSF can be determined. An elevated opening pressure and improvement in headache following removal of CSF are diagnostic in the absence of fundal changes.

Initial treatment is with acetazolamide (250–500 mg bid); the headache may improve within weeks. If ineffective, topiramate is the next treatment of choice; it has many actions that may be useful in this setting, including carbonic anhydrase inhibition, weight loss, and neuronal membrane stabilization, likely mediated via effects on phosphorylation pathways. Severely disabled patients who do not respond to medical treatment require intracranial pressure monitoring and may require shunting. If appropriate, weight loss should be encouraged.

Posttraumatic Headache

A traumatic event can trigger a headache process that lasts for many months or years after the event. The term *trauma* is used here in a very broad sense: headache can develop following an injury to the head, but it can also develop after an infectious episode, typically viral meningitis; a flulike illness; or a parasitic infection. Complaints of dizziness, vertigo, and impaired memory can accompany the headache. Symptoms may remit after several weeks or persist for months and even years after the injury. Typically, the neurologic examination is normal and CT or MRI studies are unrevealing. Chronic subdural hematoma may on occasion mimic this disorder. Posttraumatic headache may also be seen after carotid dissection and subarachnoid hemorrhage and after intracranial surgery. The underlying theme appears to be that a traumatic event involving the pain-producing meninges can trigger a headache process that lasts for many years.

Other Causes

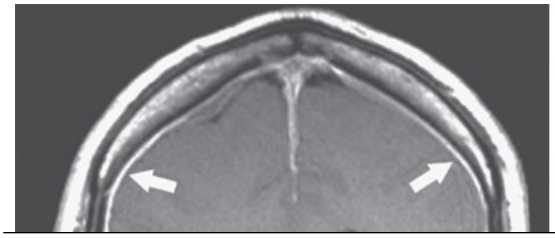
In one series, one-third of patients with NDPH reported headache beginning after a transient flulike illness characterized by fever, neck stiffness, photophobia, and marked malaise. Evaluation typically reveals no apparent cause for the headache. There is no convincing evidence that persistent Epstein-Barr virus infection plays a role in NDPH. A complicating factor is that many patients undergo LP during the acute illness; iatrogenic low CSF volume headache must be considered in these cases.

Treatment

Treatment is largely empirical and directed at the headache phenotype. Tricyclic antidepressants, notably amitriptyline, and anticonvulsants, such as topiramate, valproate, candesartan, and gabapentin, have been used with reported benefit. The monoamine oxidase inhibitor phenelzine may also be useful in carefully selected patients. The headache usually resolves within 3–5 years, but it can be quite disabling.

FIGURE 16-1

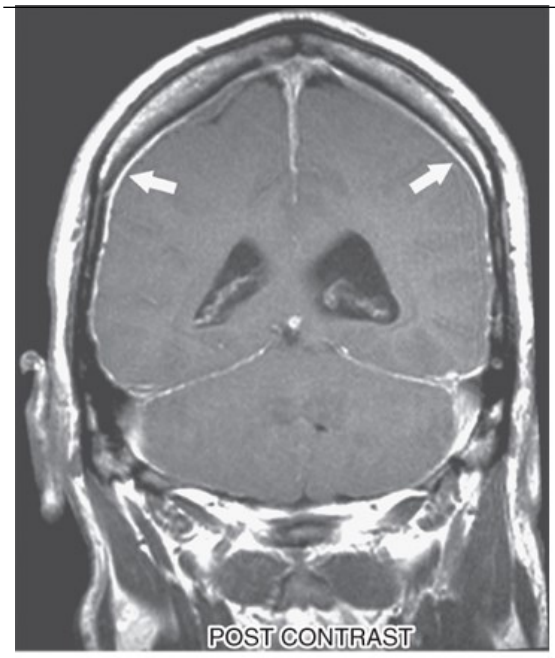
Magnetic resonance image showing diffuse meningeal enhancement after gadolinium administration in a patient with low cerebrospinal fluid (CSF) volume headache. (*Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.*)



Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Edition Copyright © McGraw Hill Education. All rights reserved.)

Access Provided by: SILVERCHAIR



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 16-4

Differential Diagnosis of New Daily Persistent Headache

PRIMARY	SECONDARY
Migrainous-type	Subarachnoid hemorrhage
Featureless (tension-type)	Low cerebrospinal fluid (CSF) volume headache
	Raised CSF pressure headache
	Posttraumatic headache[a]
	Chronic meningitis

[a]Includes postinfectious forms.

PRIMARY CARE AND HEADACHE MANAGEMENT

Most patients with headache will be seen first in a primary care setting. The challenging task of the primary care physician is to identify the very few worrisome secondary headaches from the very great majority of primary and less dangerous secondary headaches (Table 16-2).

Absent any warning signs, a reasonable approach is to treat when a diagnosis is established. As a general rule, the investigation should focus on identifying worrisome causes of headache or on helping the patient to gain confidence if no primary headache diagnosis can be made.

After treatment has been initiated, follow-up care is essential to identify whether progress has been made against the headache complaint. Not all headaches will respond to treatment, but, in general, worrisome headaches will progress and will be easier to identify.

When a primary care physician feels the diagnosis is a primary headache disorder, it is worth noting that >90% of patients who present to primary care unresponsive to two or more standard therapies for the considered headache type, should be considered for referral to a specialist. In a practical sense, the threshold for referral is also determined by the experience of the primary care physician in headache medicine and the availability of

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Absent any warning signs, a reasonable approach is to treat when a diagnosis is established. As a general rule, the investigation should focus on headaches will respond to treatment, but, in general, worrisome headaches will progress and will be easier to identify.

When a primary care physician feels the diagnosis is a primary headache disorder, it is worth noting that >90% of patients who present to primary care with a complaint of headache will have migraine **(Chap. 430)**.

In general, patients who do not have a clear diagnosis, have a primary headache disorder other than migraine or tension-type headache, or are unresponsive to two or more standard therapies for the considered headache type, should be considered for referral to a specialist. In a practical sense, the threshold for referral is also determined by the experience of the primary care physician in headache medicine and the availability of secondary care options.

ACKNOWLEDGMENT

The editors acknowledge the contributions of Neil H. Raskin to earlier editions of this chapter.

FURTHER READING

Headache Classification Committee of the International Headache Society: The International Classification of Headache Disorders, 3rd ed. Cephalalgia 33:629, 2018.

Kernick D, Goadsby PJ: *Headache: A Practical Manual*. Oxford: Oxford University Press, 2008.

Lance JW, Goadsby PJ: *Mechanism and Management of Headache*, 7th ed. New York, Elsevier, 2005.

Olesen J et al: *The Headaches*. Philadelphia, Lippincott, Williams & Wilkins, 2005.

Silberstein SD, Lipton RB, Dodick DW: *Wolff's Headache and Other Head Pain*, 9th ed. New York, Oxford University Press, 2021.

Downloaded 2022-1-3 5:44 P Your IP is 141.216.78.40
Chapter 16: Headache, Peter J. Goadsby
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:

Chapter 17: Back and Neck Pain

John W. Engstrom

INTRODUCTION

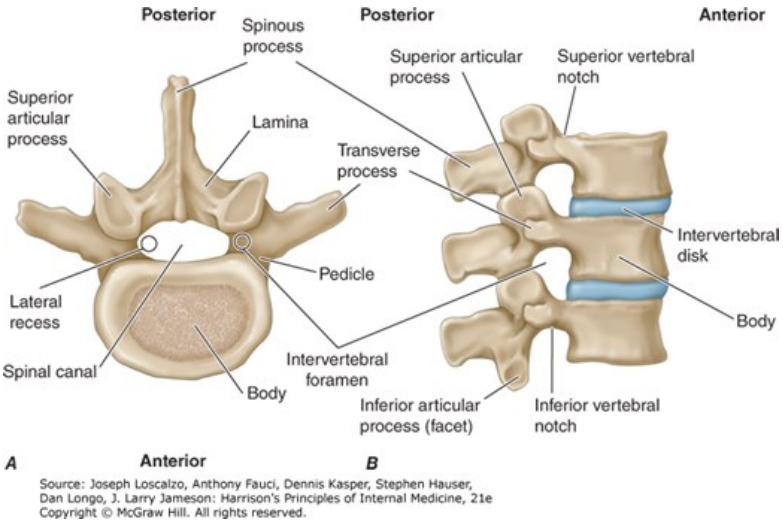
The importance of back and neck pain in our society is underscored by the following: (1) the cost of chronic back pain in the United States is estimated at more than $200 billion annually; approximately one-third of this cost is due to direct health care expenses and two-thirds are indirect costs resulting from loss of wages and productivity; (2) back symptoms are the most common cause of disability in individuals <45 years of age; (3) low back pain (LBP) is the second most common reason for visiting a physician in the United States; and (4) more than four out of five people will experience significant back pain at some point in their lives.

ANATOMY OF THE SPINE

The anterior spine consists of cylindrical vertebral bodies separated by intervertebral disks and stabilized by the anterior and posterior longitudinal ligaments. The intervertebral disks are composed of a central gelatinous nucleus pulposus surrounded by a tough cartilaginous ring, the annulus fibrosis. Disks are responsible for 25% of spinal column length and allow the bony vertebrae to move easily upon each other **(Figs. 17-1 and 17-2)**. Desiccation of the nucleus pulposus and degeneration of the annulus fibrosus worsen with age, resulting in loss of disk height. The disks are largest in the cervical and lumbar regions where movements of the spine are greatest. The anterior spine absorbs the shock of bodily movements such as walking and running, and with the posterior spine protects the spinal cord and nerve roots in the spinal canal.

FIGURE 17-1

Vertebral anatomy. ***A.*** Vertebral body—axial view; ***B.*** vertebral column—sagittal view. *(Reproduced with permission from A Gauthier Cornuelle, DH Gronefeld: Radiographic Anatomy Positioning. New York, McGraw-Hill, 1998.)*



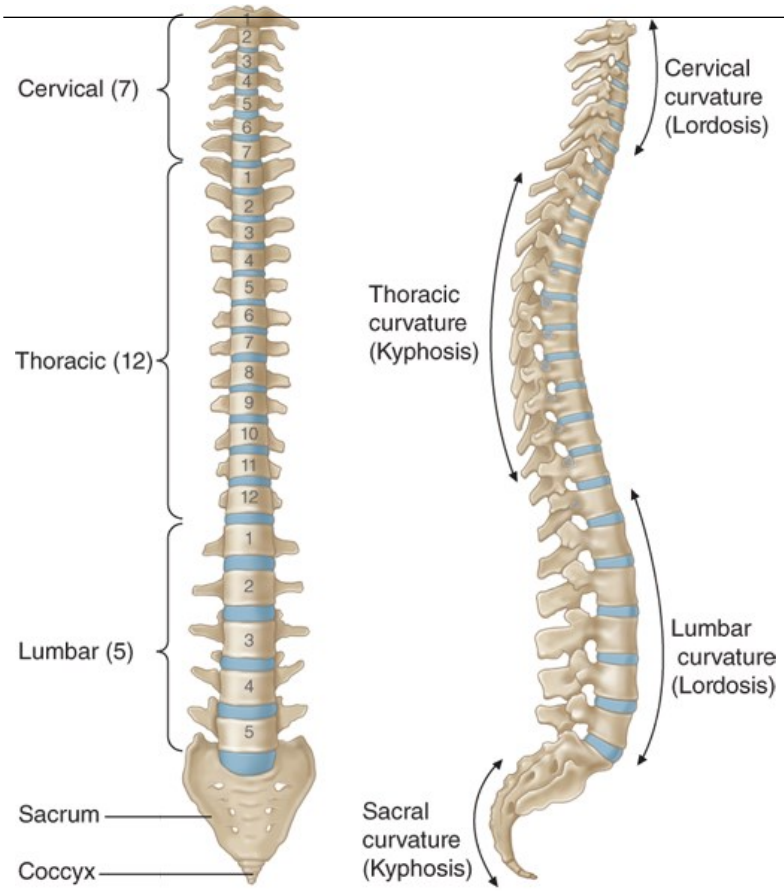
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 17-2

Spinal column. *(Reproduced with permission from A Gauthier Cornuelle, DH Gronefeld: Radiographic Anatomy Positioning. New York, McGraw-Hill, 1998.)*

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The posterior spine consists of the vertebral arches and processes. Each arch consists of paired cylindrical pedicles anteriorly and paired lamina posteriorly. The vertebral arch also gives rise to two transverse processes laterally, one spinous process posteriorly, plus two superior and two inferior articular facets. The apposition of a superior and inferior facet constitutes a *facet joint*. The posterior spine provides an anchor for the attachment of muscles and ligaments. The contraction of muscles attached to the spinous and transverse processes and lamina works like a system of pulleys and levers producing flexion, extension, rotation, and lateral bending movements of the spine.

Nerve root injury (*radiculopathy*) is a common cause of pain in the neck and arm, or low back and buttock, or leg **(see dermatomes in Figs. 25-2 and 25-3)**. Each nerve root exits just above its corresponding vertebral body in the cervical region (e.g., the C7 nerve root exits at the C6-C7 level), and just below the vertebral body in the thoracic and lumbar spine (e.g., the T1 nerve root exits at the T1-T2 level). The cervical nerve roots follow a short intraspinal course before exiting. In contrast, because the spinal cord ends at the L1 or L2 vertebral level, the lumbar nerve roots follow a long intraspinal course and can be injured anywhere along its path. For example, disk herniation at the L4-L5 level can produce L4 root compression laterally, but more often compression of the traversing L5 nerve root occurs **(Fig. 17-3)**. The lumbar nerve roots are mobile in the spinal canal, but eventually pass through the narrow *lateral recess* of the spinal canal and *intervertebral foramen* (Figs. 17-2 and 17-3). When imaging the spine, both sagittal and axial views are needed to assess possible compression at these sites.

FIGURE 17-3

Compression of L5 and S1 roots by herniated disks. *(Reproduced with permission from AH Ropper, MA Samuels: Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009.)*

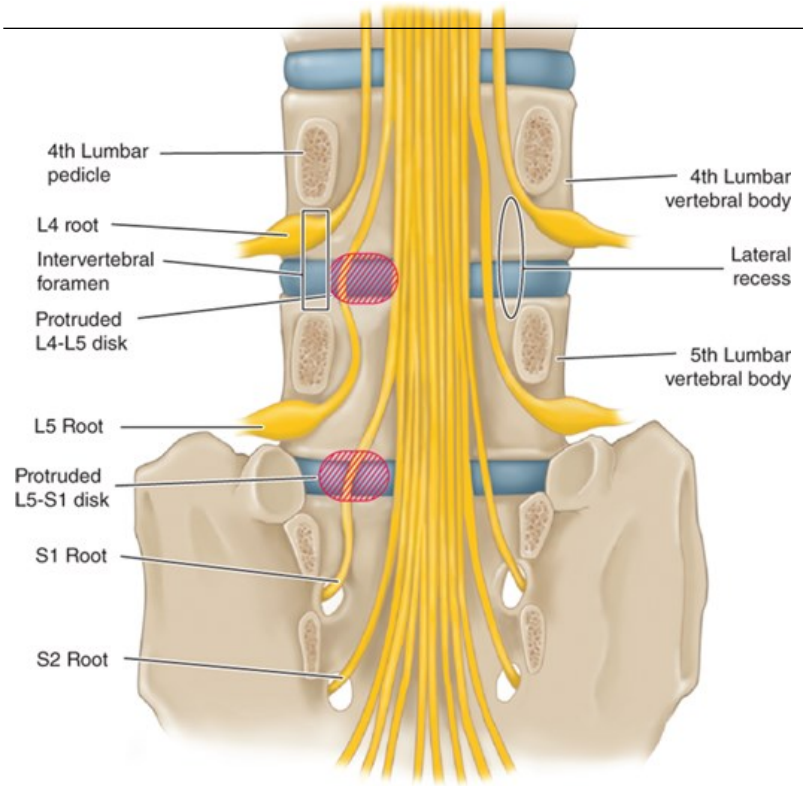


Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



FIGURE 17-3

Access Provided by:



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Beginning at the C3 level, each cervical (and the first thoracic) vertebral body projects a lateral bony process upward—the uncinate process. The uncinate process articulates with the cervical vertebral body above via the uncovertebral joint. The uncovertebral joint can hypertrophy with age and contribute to neural foraminal narrowing and cervical radiculopathy.

Pain-sensitive structures of the spine include the periosteum of the vertebrae, dura, facet joints, annulus fibrosus of the intervertebral disk, epidural veins and arteries, and the longitudinal ligaments. Disease of these diverse structures may explain many cases of back pain without nerve root compression. Under normal circumstances, the nucleus pulposus of the intervertebral disk is not pain sensitive.

APPROACH TO THE PATIENT

Back Pain

TYPES OF BACK PAIN

Delineating the type of pain reported by the patient is the essential first step. Attention is also focused on identifying risk factors for a serious underlying etiology. The most frequent serious causes of back pain are radiculopathy, fracture, tumor, infection, or referred pain from visceral structures **(Table 17-1)**.

Local pain is caused by injury to pain-sensitive structures that compress or irritate sensory nerve endings. The site of the pain is near the affected part of the back.

Pain referred to the back may arise from abdominal or pelvic viscera. The pain is usually described as primarily abdominal or pelvic, accompanied by back pain, and usually unaffected by posture. The patient may occasionally complain of back pain only.

Pain of spine origin may be located in the back or referred to the buttocks or legs. Diseases affecting the upper lumbar spine tend to refer pain to the lumbar region, groin, or anterior thighs. Diseases affecting the lower lumbar spine tend to produce pain referred to the buttocks, posterior thighs, calves, or feet. Referred pain often explains pain syndromes that cross multiple dermatomes without evidence of nerve or nerve root injury.

Radicular pain is typically sharp and radiates from the low back to a leg within the territory of a nerve root (see "Lumbar Disk Disease," below). Coughing, sneezing, or voluntary contraction of abdominal muscles (lifting heavy objects or straining at stool) may elicit or worsen the radiating and is not stretched by sitting. The description of the pain alone often fails to distinguish between referred pain and radiculopathy, although a burning or electric quality favors radiculopathy.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:

Radicular pain is typically sharp and radiates from the low back to a leg within the territory of a nerve root (see "Lumbar Disk Disease," below). Coughing, sneezing, or voluntary contraction of abdominal muscles (lifting heavy objects or straining at stool) may elicit or worsen the radiating pain. The pain may also increase in postures that stretch the nerves and nerve roots. Sitting with the leg outstretched places traction on the sciatic nerve and L5 and S1 roots because the sciatic nerve passes posterior to the hip. The femoral nerve (L2, L3, and L4 roots) passes anterior to the hip and is not stretched by sitting. The description of the pain alone often fails to distinguish between referred pain and radiculopathy, although a burning or electric quality favors radiculopathy.

Pain associated with muscle spasm is commonly associated with many spine disorders. The spasms may be accompanied by an abnormal posture, tense paraspinal muscles, and dull or achy pain in the paraspinal region.

Knowledge of the circumstances associated with the onset of back pain is important when weighing possible serious underlying causes for the pain. Some patients involved in accidents or work-related injuries may exaggerate their pain for the purpose of compensation or for psychological reasons.

EXAMINATION

A complete physical examination including vital signs, heart and lungs, abdomen and rectum, and limbs is advisable. Back pain referred from visceral organs may be reproduced during palpation of the abdomen (pancreatitis, abdominal aortic aneurysm [AAA]) or percussion over the costovertebral angles (pyelonephritis).

The normal spine has a cervical and lumbar lordosis and a thoracic kyphosis. Exaggeration of these normal alignments may result in hyperkyphosis of the thoracic spine or hyperlordosis of the lumbar spine. Inspection of the back may reveal a lateral curvature of the spine (scoliosis). A midline hair tuft, skin dimpling or pigmentation, or a sinus tract may indicate a congenital spine anomaly. Asymmetry in the prominence of the paraspinal muscles suggests muscle spasm. Palpation over the spinous process transmits force to the entire vertebrae and suggests vertebral pathology.

Flexion at the hips is normal in patients with lumbar spine disease, but flexion of the lumbar spine is limited and sometimes painful. Lateral bending to the side opposite the injured spinal element may stretch the damaged tissues, worsen pain, and limit motion. Hyperextension of the spine (with the patient prone or standing) is limited when nerve root compression, facet joint pathology, or other bony spine disease is present.

Pain from hip disease may mimic the pain of lumbar spine disease. Hip pain can be reproduced by passive internal and external rotation at the hip with the knee and hip in flexion or by percussing the heel with the examiner's palm with the leg extended (heel percussion sign).

The *straight-leg raising (SLR)* maneuver is a simple bedside test for nerve root disease. With the patient supine, passive straight-leg flexion at the hip stretches the L5 and S1 nerve roots and the sciatic nerve; dorsiflexion of the foot during the maneuver adds to the stretch. In healthy individuals, flexion to at least 80° is normally possible without causing pain, although a tight, stretching sensation in the hamstring muscles is common. The SLR test is positive if the maneuver reproduces the patient's usual back or limb pain. Eliciting the SLR sign in both the supine and sitting positions can help determine if the finding is reproducible. The patient may describe pain in the low back, buttocks, posterior thigh, or lower leg, but the *key feature is reproduction of the patient's usual pain.* The *crossed SLR sign* is present when flexion of one leg reproduces the usual pain in the opposite leg or buttocks. In disk herniation, the crossed SLR sign is less sensitive but more specific than the SLR sign. The *reverse SLR sign* is elicited by standing the patient next to the examination table and passively extending each leg with the knee fully extended. This maneuver, which stretches the L2-L4 nerve roots, lumbosacral plexus, and femoral nerve, is considered positive if the patient's usual back or limb pain is reproduced. For all of these tests, the nerve or nerve root lesion is always on the side of the pain. Examination of the unaffected leg first provides a control test, ensures mutual understanding of test parameters, and enhances test utility.

The neurologic examination includes a search for focal weakness or muscle atrophy, localized reflex changes, diminished sensation in the legs, or signs of spinal cord injury. The examiner should be alert to the possibility of breakaway weakness, defined as fluctuations in the maximum power generated during muscle testing. Breakaway weakness may be due to pain, inattention, or a combination of pain and underlying true weakness. Breakaway weakness without pain is usually due to a lack of effort. In uncertain cases, electromyography (EMG) can determine if true weakness due to nerve tissue injury is present. Findings with specific lumbosacral nerve root lesions are shown in **Table 17-2** and are discussed below.

LABORATORY, IMAGING, AND EMG STUDIES

Laboratory studies are rarely needed for the initial evaluation of nonspecific acute (<3 months duration) low back pain (ALBP). Risk factors for a serious underlying cause and for infection, tumor, or fracture in particular should be sought by history and examination. If risk factors are present (Table 17-1), then laboratory studies (complete blood count [CBC], erythrocyte sedimentation rate [ESR], urinalysis) are indicated. If risk factors are

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

involving posterior spine structures, craniocervical and cervicothoracic junctions, C1 and C2 vertebrae, bone fragments in the spinal canal, or misalignment. MRI or CT myelography is the radiologic test of choice for evaluation of most serious diseases involving the spine. MRI is superior for

Access Provided by:

serious underlying cause and for infection, tumor, or fracture in particular should be sought by history and examination. If risk factors are present (Table 17-1), then laboratory studies (complete blood count [CBC], erythrocyte sedimentation rate [ESR], urinalysis) are indicated. If risk factors are absent, then management is conservative (see "Treatment," below).

CT scanning is used as a primary screening modality for acute trauma that is moderate to severe. CT is superior to x-rays for detection of fractures involving posterior spine structures, craniocervical and cervicothoracic junctions, C1 and C2 vertebrae, bone fragments in the spinal canal, or misalignment. MRI or CT myelography is the radiologic test of choice for evaluation of most serious diseases involving the spine. MRI is superior for the definition of soft tissue structures, whereas CT myelography provides optimal imaging of the lateral recess of the spinal canal, defines bony abnormalities, and is tolerated by claustrophobic patients.

Population surveys in the United States suggest that patients with back pain report greater functional limitations in recent years, despite rapid increases in spine imaging, opioid prescribing, injections, and spine surgery. This suggests that more selective use of diagnostic and treatment modalities may be reasonable for many patients. One prospective case-control study found that older adults with back pain of less than 6 weeks duration who received spine imaging as part of a primary care visit had no better outcomes than the control group.

Spine imaging often reveals abnormalities of dubious clinical relevance that may alarm clinicians and patients alike and prompt further testing and unnecessary therapy. When imaging tests are reviewed, it is important to remember that degenerative findings are common in normal, pain-free individuals. Randomized trials and observational studies have suggested that imaging can have a "cascade effect," creating a gateway to other unnecessary care. Interventions have included physician education and computerized decision support within the electronic medical record to require specific indications for approval of imaging tests. Other strategies have included audit and feedback of individual practitioners' rates of ordering, more rapid access to physical therapy, or consultation with spine experts for patients without imaging indications.

Educational tools created by the America College of Physicians for patients and the public have included "Five Things Physicians and Patients Should Question": (1) Do not recommend advanced imaging (e.g., MRI) of the spine within the first 6 weeks in patients with nonspecific ALBP in the absence of red flags. (2) Do not perform elective spinal injections without imaging guidance, unless contraindicated. (3) Do not use bone morphogenetic protein (BMP) for routine anterior cervical spine fusion surgery. (4) Do not use EMG and nerve conduction studies (NCSs) to determine the cause of purely midline lumbar, thoracic, or cervical spine pain. (5) Do not recommend bed rest for >48 h when treating LBP. In an observational study, application of this strategy was associated with lower rates of repeat imaging, opioid use, and referrals for physical therapy.

Electrodiagnostic studies can be used to assess the functional integrity of the peripheral nervous system **(Chap. 446)**. Sensory NCSs are normal when focal sensory loss confirmed by examination is due to nerve root damage because the nerve roots are proximal to the nerve cell bodies in the dorsal root ganglia. Injury to nerve tissue distal to the dorsal root ganglion (e.g., plexus or peripheral nerve) results in reduced sensory nerve signals. Needle EMG complements NCSs by detecting denervation or reinnervation changes in a myotomal (segmental) distribution. Multiple muscles supplied by different nerve roots and nerves are sampled; the pattern of muscle involvement indicates the nerve root(s) responsible for the injury. Needle EMG provides objective information about motor nerve fiber injury when clinical evaluation of weakness is limited by pain or poor effort. EMG and NCSs will be normal when sensory nerve root injury or irritation is the pain source.

The COVID-19 pandemic has disrupted and complicated the care of patients with LBP. Paraspinal myalgias may result in LBP. The sedentary lifestyle resulting from quarantine is associated with an increased frequency or severity of LBP. Fear of infection risk has also prevented many patients from seeking needed care. Video-telemedicine visits can help identify patients with underlying risks for a serious cause and inform appropriate next steps in management.

TABLE 17-1

Acute Low Back Pain: Risk Factors for an Important Structural Cause

History
Pain worse at rest or at night
Prior history of cancer
History of chronic infection (especially lung, urinary tract, skin, poor dentition)
Incontinence

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:

TABLE 17-1

Acute Low Back Pain: Risk Factors for an Important Structural Cause

History
Pain worse at rest or at night
Prior history of cancer
History of chronic infection (especially lung, urinary tract, skin, poor dentition)
History of trauma
Incontinence
Age >70 years
Intravenous drug use
Glucocorticoid use
History of a rapidly progressive neurologic deficit
Examination
Unexplained fever
Unexplained weight loss
Focal palpation/percussion tenderness over the midline spine
Abdominal, rectal, or pelvic mass
Internal/external rotation of the leg at the hip
Straight-leg or reverse straight-leg raising signs
Progressive focal neurologic deficit

TABLE 17-2

Lumbosacral Radiculopathy: Neurologic Features

LUMBOSACRAL NERVE ROOT	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
L2[a]	—	Upper anterior thigh	Psoas (hip flexors)	Anterior thigh
L3[a]	—	Lower anterior thigh	Psoas (hip flexors)	Anterior thigh, knee
			Thigh adductors	

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:

TABLE 17-2

Lumbosacral Radiculopathy: Neurologic Features

LUMBOSACRAL NERVE ROOT	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
L2[a]	—	Upper anterior thigh	Psoas (hip flexors)	Anterior thigh
L3[a]	—	Lower anterior thigh	Psoas (hip flexors)	Anterior thigh, knee
		Anterior knee	Quadriceps (knee extensors)	
			Thigh adductors	
L4[a]	Quadriceps (knee)	Medial calf	Quadriceps (knee extensors)[b]	Knee, medial calf
			Thigh adductors	Anterolateral thigh
L5[c]	—	Dorsal surface—foot	Peronei (foot evertors)[b]	Lateral calf, dorsal foot, posterolateral thigh, buttocks
		Lateral calf	Tibialis anterior (foot dorsiflexors)	
			Gluteus medius (leg abductors)	
			Toe dorsiflexors	
S1[c]	Gastrocnemius/soleus (ankle)	Plantar surface—foot	Gastrocnemius/soleus (foot plantar flexors)[b]	Bottom foot, posterior calf, posterior thigh, buttocks
		Lateral aspect—foot	Abductor hallucis (toe flexors)[b]	
			Gluteus maximus (leg extensors)	

[a]Reverse straight-leg raising sign may be present—see “Examination of the Back.”

[b]These muscles receive the majority of innervation from this root.

[c]Straight-leg raising sign may be present—see “Examination of the Back.”

CAUSES OF BACK PAIN

(TABLE 17-3)

TABLE 17-3

Causes of Back or Neck Pain

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Degenerative Spine Disease

Access Provided by:

Causes of Back or Neck Pain

Lumbar or Cervical Disk Disease
Degenerative Spine Disease
Lumbar spinal stenosis without or with neurogenic claudication Intervertebral foraminal or lateral recess narrowing Disk-osteophyte complex Facet or uncovertebral joint hypertrophy Lateral disk protrusion Spondylosis (osteoarthritis), spondylolisthesis, or spondylolysis
Spine Infection
Vertebral osteomyelitis Spinal epidural abscess Septic disk (diskitis) Meningitis Lumbar arachnoiditis
Neoplasms
Metastatic with/without pathologic fracture Primary Nervous System: Meningioma, neurofibroma, schwannoma Primary Bone: chordoma, osteoma
Trauma
Strain or sprain Whiplash injury Trauma/falls, motor vehicle accidents
Metabolic Spine Disease
Osteoporosis with/without pathologic fracture—hyperparathyroidism, immobility Osteosclerosis (e.g., Paget's disease)
Congenital/Developmental
Spondylolysis Kyphoscoliosis Spina bifida occulta Tethered spinal cord
Autoimmune Inflammatory Arthritis
Other Causes of Back Pain Referred pain from visceral disease (e.g., abdominal aortic aneurysm) Postural Psychiatric, malingering, chronic pain syndromes

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

LUMBAR DISK DISEASE

Access Provided by:

Postural

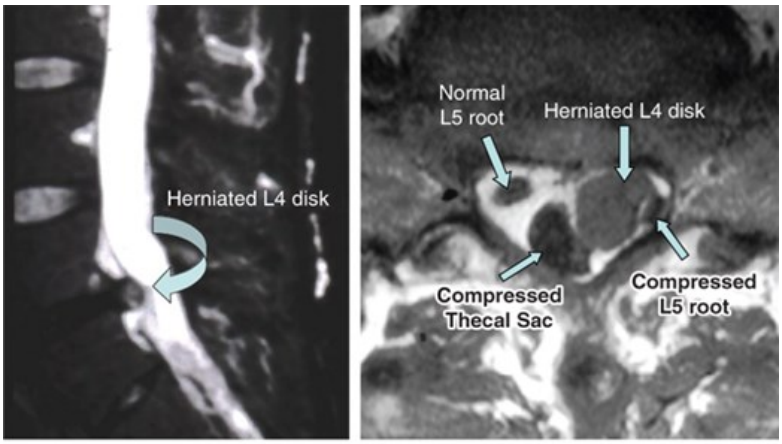
Psychiatric, malingering, chronic pain syndromes

LUMBAR DISK DISEASE

Lumbar disk disease is a common cause of acute, chronic, or recurrent low back and leg pain **(Figs. 17-3 and 17-4)**. Disk disease is most likely to occur at the L4-L5 or L5-S1 levels, but upper lumbar levels can also be involved. The cause is often unknown, but the risk is increased in overweight individuals. Disk herniation is unusual prior to age 20 years and is rare in the fibrotic disks of the elderly. Complex genetic factors may play a role in predisposition. The pain may be located in the low back only or referred to a leg, buttock, or hip. A sneeze, cough, or trivial movement may cause the nucleus pulposus to prolapse, pushing the frayed and weakened annulus posteriorly. With severe disk disease, the nucleus can protrude through the annulus (herniation) or become extruded to lie as a free fragment in the spinal canal.

FIGURE 17-4

Disk herniation. ***A***. Sagittal T2-weighted image on the left side of the spinal canal reveals disk herniation at the L4-L5 level. ***B***. Axial T1-weighted image shows paracentral disk herniation with displacement of the thecal sac medially and the left L5 nerve root posteriorly in the left lateral recess. (*Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.*)



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The mechanism by which intervertebral disk injury causes back pain is uncertain. The inner annulus fibrosus and nucleus pulposus are normally devoid of innervation. Inflammation and production of proinflammatory cytokines within a ruptured nucleus pulposus may trigger or perpetuate back pain. Ingrowth of nociceptive (pain) nerve fibers into the nucleus pulposus of a diseased disk may be responsible for some cases of chronic "diskogenic" pain. Nerve root injury (radiculopathy) from disk herniation is usually due to inflammation, but lateral herniation may produce compression in the lateral recess or intervertebral foramen.

A ruptured disk may be asymptomatic or cause back pain, limited spine motion (particularly flexion), a focal neurologic deficit, or radicular pain. A dermatomal pattern of sensory loss or a reduced or absent deep tendon reflex is more suggestive of a specific root lesion than is the pattern of pain. Motor findings (focal weakness, muscle atrophy, or fasciculations) occur less frequently than focal sensory or reflex changes. Symptoms and signs are usually unilateral, but bilateral involvement does occur with large central disk herniations that involve roots bilaterally or cause inflammation of nerve roots within the spinal canal. Clinical manifestations of specific nerve root lesions are summarized in Table 17-2.

The differential diagnosis covers a variety of serious and treatable conditions, including epidural abscess, hematoma, fracture, or tumor. Fever, constant pain uninfluenced by position, sphincter abnormalities, or signs of myelopathy suggest an etiology other than lumbar disk disease. Absent ankle reflexes can be a normal finding in persons >60 years or a sign of bilateral S1 radiculopathies. An absent deep tendon reflex or focal sensory loss may indicate injury to a nerve root, but other sites of injury along the nerve must also be considered. As examples, an absent knee reflex may be due to a femoral neuropathy or an L4 nerve root injury; loss of sensation over the foot and lateral lower calf may result from a peroneal or lateral sciatic neuropathy or an L5 nerve root injury. Focal muscle atrophy may reflect injury to the anterior horn cells of the spinal cord, a nerve root, peripheral

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

adjacent soft tissue anatomy, whereas bony lesions of the lateral recess or intervertebral foramen are optimally visualized by CT myelography. The correlation of neuroradiologic findings to clinical symptoms, particularly pain, is not simple. Contrast-enhancing tears in the annulus fibrosus or disk

Access Provided by:

a femoral neuropathy or an L4 nerve root injury; loss of sensation over the foot and lateral lower calf may result from a peroneal or lateral sciatic neuropathy, or an L5 nerve root injury. Focal muscle atrophy may reflect injury to the anterior horn cells of the spinal cord, a nerve root, peripheral nerve, or disuse.

A lumbar spine MRI scan or CT myelogram can often confirm the location and type of pathology. Spine MRIs yield exquisite views of intraspinal and adjacent soft tissue anatomy, whereas bony lesions of the lateral recess or intervertebral foramen are optimally visualized by CT myelography. The correlation of neuroradiologic findings to clinical symptoms, particularly pain, is not simple. Contrast-enhancing tears in the annulus fibrosus or disk protrusions are widely accepted as common sources of back pain; however, studies have found that many asymptomatic adults have similar radiologic findings. Entirely asymptomatic disk protrusions are also common, occurring in up to one-third of adults, and these may also enhance with contrast. Furthermore, in patients with known disk herniation treated either medically or surgically, persistence of the herniation 10 years later had no relationship to the clinical outcome. In summary, MRI findings of disk protrusion, tears in the annulus fibrosus, or hypertrophic facet joints are common incidental findings that, by themselves, should not dictate management decisions for patients with back pain.

The diagnosis of nerve root injury is most secure when the history, examination, results of imaging studies, and the EMG are concordant. There is often good correlation between CT and EMG findings for localization of nerve root injury.

Management of lumbar disk disease is discussed below. *Cauda equina syndrome* (*CES*) signifies an injury of multiple lumbosacral nerve roots within the spinal canal distal to the termination of the spinal cord at L1-L2. LBP, weakness and areflexia in the legs, saddle anesthesia, or loss of bladder function may occur. The problem must be distinguished from disorders of the lower spinal cord (conus medullaris syndrome), acute transverse myelitis **(Chap. 442)**, and Guillain-Barré syndrome **(Chap. 447)**. Combined involvement of the conus medullaris and cauda equina can occur. CES is most commonly due to a large ruptured lumbosacral intervertebral disk, but other causes include lumbosacral spine fracture, hematoma within the spinal canal (sometimes following lumbar puncture in patients with coagulopathy), and tumor or other compressive mass lesions. Treatment is usually surgical decompression, sometimes on an urgent basis in an attempt to restore or preserve motor or sphincter function, or radiotherapy for metastatic tumors **(Chap. 90)**.

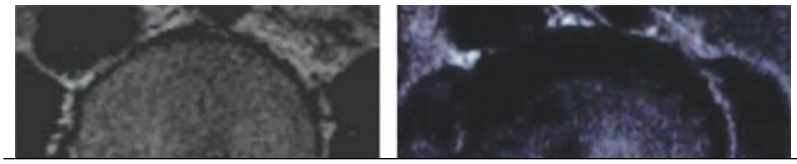
DEGENERATIVE CONDITIONS

Lumbar spinal stenosis (*LSS*) describes a narrowed lumbar spinal canal. *Neurogenic claudication* consists of pain, typically in the back and buttocks or legs, that is brought on by walking or standing and relieved by sitting. Unlike vascular claudication, symptoms are often provoked by standing without walking. Unlike lumbar disk disease, symptoms are usually relieved by sitting. Patients with neurogenic claudication can often walk much farther when leaning over a shopping cart and can pedal a stationary bike with ease while sitting. These flexed positions increase the anteroposterior spinal canal diameter and reduce intraspinal venous hypertension, producing pain relief. Focal weakness, sensory loss, or reflex changes may occur when spinal stenosis is associated with neural foraminal narrowing and radiculopathy. Severe neurologic deficits, including paralysis and urinary incontinence, occur only rarely.

LSS by itself is common (6–7% of adults) and is usually asymptomatic. Symptoms are correlated with severe spinal canal stenosis. LSS is most often acquired (75%) but can also be congenital or due to a mixture of both etiologies. Congenital forms (achondroplasia and idiopathic) are characterized by short, thick pedicles that produce both spinal canal and lateral recess stenosis. Acquired factors that contribute to spinal stenosis include degenerative diseases (spondylosis, spondylolisthesis, and scoliosis), trauma, spine surgery, metabolic or endocrine disorders (epidural lipomatosis, osteoporosis, acromegaly, renal osteodystrophy, and hypoparathyroidism), and Paget's disease. MRI provides the best definition of the abnormal anatomy **(Fig. 17-5)**.

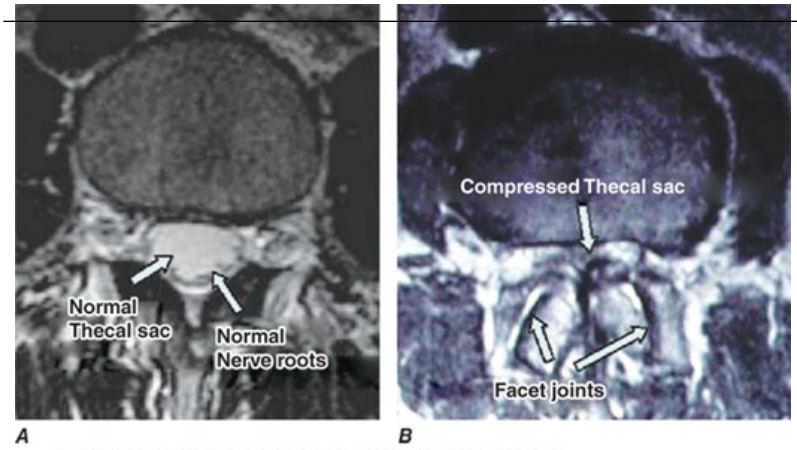
FIGURE 17-5

Spinal stenosis. ***A***. An axial T2-weighted image of the normal lumbar spine shows a normal thecal sac within the lumbar spinal canal. The thecal sac is bright. The lumbar roots are seen as dark punctate dots located posteriorly in the thecal sac. ***B***. The thecal sac is not well visualized due to severe lumbar spinal canal stenosis, partially the result of hypertrophic facet joints. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

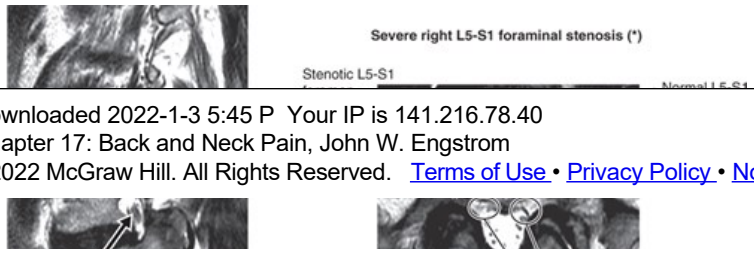
LSS accompanied by neurogenic claudication responds to surgical decompression of the stenotic segments. The same processes leading to LSS may cause lumbar foraminal or lateral recess narrowing resulting in coincident lumbar radiculopathy that may require treatment as well.

Conservative treatment of symptomatic LSS can include nonsteroidal anti-inflammatory drugs (NSAIDs), acetaminophen, exercise programs, and symptomatic treatment of acute pain episodes. There is insufficient evidence to support the routine use of epidural glucocorticoid injections. Surgery is considered when medical therapy does not relieve symptoms sufficiently to allow for resumption of activities of daily living or when focal neurologic signs are present. Most patients with neurogenic claudication who are treated medically do not improve over time. Surgical management with laminectomy, which increases the spinal canal diameter and reduces venous hypertension, can produce significant relief of exertional back and leg pain, leading to less disability and improved functional outcomes. Laminectomy and fusion is usually reserved for patients with LSS and spondylolisthesis. Predictors of a poor surgical outcome include impaired walking preoperatively, depression, cardiovascular disease, and scoliosis. Up to one-quarter of surgically treated patients develop recurrent stenosis at the same or an adjacent spinal level within 7–10 years; recurrent symptoms usually respond to a second surgical decompression.

Neural foraminal narrowing or lateral recess stenosis with radiculopathy is a common consequence of osteoarthritic processes that cause LSS **(Figs. 17-1 and 17-6)**, including osteophytes, lateral disk protrusion, calcified disk-osteophytes, facet joint hypertrophy, uncovertebral joint hypertrophy (in the cervical spine), congenitally shortened pedicles, or, frequently, a combination of these processes. Neoplasms (primary or metastatic), fractures, infections (epidural abscess), or hematomas are less frequent causes. Most common is bony foraminal narrowing leading to nerve root ischemia and persistent symptoms, in contrast to inflammation that is associated with a paracentral herniated disk and radiculopathy. These conditions can produce unilateral nerve root symptoms or signs due to compression at the intervertebral foramen or in the lateral recess; symptoms are indistinguishable from disk-related radiculopathy, but treatment may differ depending on the etiology. The history and neurologic examination alone cannot distinguish between these possibilities. Neuroimaging (CT or MRI) is required to identify the anatomic cause. Neurologic findings from the examination and EMG can help direct the attention of the radiologist to specific nerve roots, especially on axial images. For *facet joint hypertrophy with foraminal stenosis,* surgical foraminotomy produces long-term relief of leg and back pain in 80–90% of patients. Facet joint or medial branch blocks for back or neck pain are sometimes used to help determine the anatomic origin of back pain or for treatment, but there is a lack of clinical data to support their utility. Medical causes of lumbar or cervical radiculopathy unrelated to primary spine disease include infections (e.g., herpes zoster and Lyme disease), carcinomatous meningitis, diabetes, and root avulsion or traction (trauma).

FIGURE 17-6

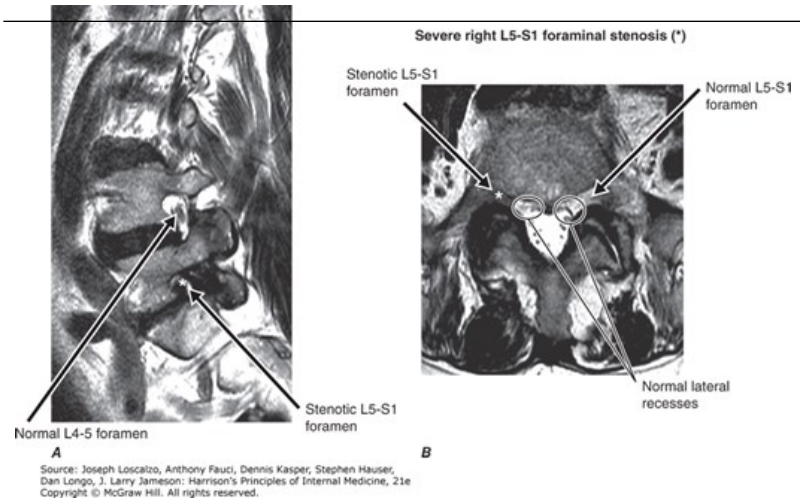
Foraminal stenosis. ***A***. Sagittal T2-weighted image reveals normal high signal around the exiting right L4 nerve root in the right neural foramen at L4-L5; effacement of the high signal is noted one level below at L5-S1, due to severe foraminal stenosis. ***B***. Axial T2-weighted image at the L5-S1 level demonstrates normal lateral recesses bilaterally, a normal intervertebral foramen on the left, but a severely stenotic foramen (*) on the right. (*Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

L4-L5; effacement of the high signal is noted one level below at L5-S1, due to severe foraminal stenosis. *B*, Axial T2-weighted image at the L5-S1 level

Access Provided by:



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

SPONDYLOSIS AND SPONDYLOLISTHESIS

Spondylosis, or osteoarthritic spine disease, typically occurs in later life and primarily involves the cervical and lumbosacral spine. Patients often complain of back pain that increases with movement, is associated with stiffness, and is better with inactivity. The relationship between clinical symptoms and radiologic findings is usually not straightforward. Pain may be prominent when MRI, CT, or x-ray findings are minimal, and prominent degenerative spine disease can be seen in asymptomatic patients. Osteophytes, combined disk-osteophytes, or a thickened ligamentum flavum may cause or contribute to central spinal canal stenosis, lateral recess stenosis, or neural foraminal narrowing.

Spondylolisthesis is the anterior slippage of the vertebral body, pedicles, and superior articular facets, leaving the posterior elements behind. Spondylolisthesis can be associated with spondylolysis, congenital anomalies, degenerative spine disease, or other causes of mechanical weakness of the pars interarticularis (e.g., infection, osteoporosis, tumor, trauma, earlier surgery). The slippage may be asymptomatic or may cause LBP, nerve root injury (the L5 root most frequently), symptomatic spinal stenosis, or CES in rare severe cases. A "step-off" on palpation or tenderness may be elicited near the segment that has "slipped" (most often L4 on L5 or occasionally L5 on S1). Focal anterolisthesis or retrolisthesis can occur at any cervical or lumbar level and be the source of neck or LBP. Plain x-rays of the low back or neck in flexion and extension will reveal movement at the abnormal spinal segment. Surgery is performed for spinal instability (slippage 5–8 mm) and considered for pain symptoms that do not respond to conservative measures (e.g., rest, physical therapy), cases with a progressive neurologic deficit, or scoliosis.

NEOPLASMS

Back pain is the most common neurologic symptom in patients with systemic cancer and is the presenting symptom in 20%. The cause is usually vertebral body metastasis (85–90%) but can also result from spread of cancer through the intervertebral foramen (especially with lymphoma), carcinomatous meningitis, or metastasis to the spinal cord. The thoracic spine is most often affected. Cancer-related back pain tends to be constant, dull, unrelieved by rest, and worse at night. By contrast, mechanical causes of LBP usually improve with rest. MRI, CT, and CT myelography are the studies of choice when spinal metastasis is suspected. Once a metastasis is found, imaging of the entire spine is essential, as it reveals additional tumor deposits in one-third of patients. MRI is preferred for soft tissue definition, but the most rapidly available imaging modality is best because the patient's condition may worsen quickly without intervention. Early diagnosis is crucial. A strong predictor of outcome is baseline neurologic function prior to diagnosis. Half to three-quarters of patients are nonambulatory at the time of diagnosis and few regain the ability to walk. **The management of spinal metastasis is discussed in detail in Chap. 90.**

INFECTIONS/INFLAMMATION

Vertebral osteomyelitis is most often caused by hematogenous seeding of staphylococci, but other bacteria or tuberculosis (Pott's disease) may be responsible. The primary source of infection is usually the skin or urinary tract. Other common sources of bacteremia are IV drug use, poor dentition, endocarditis, lung abscess, IV catheters, or postoperative wound sites. Back pain at rest, tenderness over the involved vertebra, and an elevated erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) are the most common findings in vertebral osteomyelitis. Fever or an elevated white blood cell count is found in a minority of patients. MRI and CT are sensitive and specific for early detection of osteomyelitis. The intervertebral disk can also be affected by infection (diskitis) and almost never by tumor. Extension of the infection posteriorly from the vertebral body can produce a spinal

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

signs of spinal cord compression. The subacute development of two or more of these findings should increase suspicion for spinal epidural abscess. The abscess is best delineated by spine MRI and may track over multiple spinal levels.

Access Provided by:

Vertebral osteomyelitis is most often caused by hematogenous seeding of staphylococci, but other bacteria or tuberculosis (Pott's disease) may be blood cell count is found in a minority of patients. MRI and CT are sensitive and specific for early detection of osteomyelitis. The intervertebral disk can also be affected by infection (diskitis) and almost never by tumor. Extension of the infection posteriorly from the vertebral body can produce a spinal epidural abscess.

Spinal epidural abscess **(Chap. 442)** presents with back pain (aggravated by movement or palpation of the spinous process), fever, radiculopathy, or signs of spinal cord compression. The subacute development of two or more of these findings should increase suspicion for spinal epidural abscess. The abscess is best delineated by spine MRI and may track over multiple spinal levels.

Lumbar adhesive arachnoiditis with radiculopathy is due to fibrosis following inflammation within the subarachnoid space. The fibrosis results in nerve root adhesions and presents as back and leg pain associated with multifocal motor, sensory, or reflex changes. Causes of arachnoiditis include multiple lumbar operations (most common in the United States), chronic spinal infections (especially tuberculosis in the developing world), spinal cord injury, intrathecal hemorrhage, myelography (rare), intrathecal injections (glucocorticoids, anesthetics, or other agents), and foreign bodies. The MRI shows clumped nerve roots on axial views or loculations of cerebrospinal fluid within the thecal sac. Clumped nerve roots should be distinguished from enlarged nerve roots seen with demyelinating polyneuropathy or neoplastic infiltration. Treatment is usually unsatisfactory. Microsurgical lysis of adhesions, dorsal rhizotomy, dorsal root ganglionectomy, and epidural glucocorticoids have been tried, but outcomes have been poor. Dorsal column stimulation for pain relief has produced varying results.

TRAUMA

A patient complaining of back pain and an inability to move the legs may have a spine fracture or dislocation; fractures above L1 place the spinal cord at risk for compression. Care must be taken to avoid further damage to the spinal cord or nerve roots by immobilizing the back or neck pending the results of radiologic studies. Vertebral fractures frequently occur in the absence of trauma in association with osteoporosis, glucocorticoid use, osteomyelitis, or neoplastic infiltration.

Sprains and Strains

The terms *low back sprain*, *strain*, and *mechanically induced muscle spasm* refer to minor, self-limited injuries associated with lifting a heavy object, a fall, or a sudden deceleration such as in an automobile accident. These terms are used loosely and do not correlate with specific underlying pathologies. The pain is usually confined to the lower back. Patients with paraspinal muscle spasm often assume unusual postures.

Traumatic Vertebral Fractures

Most traumatic fractures of the lumbar vertebral bodies result from injuries producing anterior wedging or compression. With severe trauma, the patient may sustain a fracture-dislocation or a "burst" fracture involving the vertebral body and posterior elements. Traumatic vertebral fractures are caused by falls from a height, sudden deceleration in an automobile accident, or direct injury. Neurologic impairment is common, and early surgical treatment is indicated. In victims of blunt trauma, CT scans of the chest, abdomen, or pelvis can be reformatted to detect associated vertebral fractures. Rules have been developed to avoid unnecessary spine imaging associated with low-risk trauma, but these studies typically exclude patients aged >65 years—a group that can sustain fractures with minor trauma.

METABOLIC CAUSES

Osteoporosis and Osteosclerosis

Immobilization, osteomalacia, the postmenopausal state, renal disease, multiple myeloma, hyperparathyroidism, hyperthyroidism, metastatic carcinoma, or glucocorticoid use may accelerate osteoporosis and weaken the vertebral body, leading to compression fractures and pain. Up to two-thirds of compression fractures seen on radiologic imaging are asymptomatic. The most common nontraumatic vertebral body fractures are due to a postmenopausal cause, or to osteoporosis in adults >75 years old **(Chap. 411)**. The risk of an additional vertebral fracture 1 year following a first vertebral fracture is 20%. The presence of fever, weight loss, fracture at a level above T4, any fracture in a young adult, or the predisposing conditions described above should increase suspicion for a cause other than typical osteoporosis. The sole manifestations of a compression fracture may be localized back or radicular pain exacerbated by movement and often reproduced by palpation over the spinous process of the affected vertebra.

Relief of acute pain can often be achieved with acetaminophen, NSAIDs, opioids, or a combination of these medications. Both pain and disability are improved with bracing. Antiresorptive drugs are not recommended in the setting of acute pain but are the preferred treatment to prevent additional fractures. Less than one-third of patients with prior compression fractures are adequately treated for osteoporosis despite the increased risk for

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Osteosclerosis, an abnormally increased bone density often due to Paget's disease, is readily identifiable on routine x-ray studies and can sometimes be a source of back pain. It may be associated with an isolated increase in alkaline phosphatase in an otherwise healthy older person. Spinal cord or

Access Provided by:

described above should increase suspicion for a cause other than typical osteoporosis. The sole manifestations of a compression fracture may be improved with bracing. Antiresorptive drugs are not recommended in the setting of acute pain but are the preferred treatment to prevent additional fractures. Less than one-third of patients with prior compression fractures are adequately treated for osteoporosis despite the increased risk for future fractures; even fewer at-risk patients without a history of fracture are adequately treated. The literature for percutaneous vertebroplasty (PVP) or kyphoplasty for osteoporotic compression fractures associated with debilitating pain does not support their use.

Osteosclerosis, an abnormally increased bone density often due to Paget's disease, is readily identifiable on routine x-ray studies and can sometimes be a source of back pain. It may be associated with an isolated increase in alkaline phosphatase in an otherwise healthy older person. Spinal cord or nerve root compression can result from bony encroachment. The diagnosis of Paget's disease as the cause of a patient's back pain is a diagnosis of exclusion.

For further discussion of these bone disorders, see Chaps. 410, 411, and 412.

AUTOIMMUNE INFLAMMATORY ARTHRITIS

Autoimmune inflammatory disease of the spine can present with the insidious onset of low back, buttock, or neck pain. Examples include rheumatoid arthritis (RA) **(Chap. 358)**, ankylosing spondylitis, reactive arthritis and psoriatic arthritis **(Chap. 355)**, or inflammatory bowel disease **(Chap. 326)**.

CONGENITAL ANOMALIES OF THE LUMBAR SPINE

Spondylolysis is a bony defect in the vertebral pars interarticularis (a segment near the junction of the pedicle with the lamina), a finding present in up to 6% of adolescents. The cause is usually a stress microfracture in a congenitally abnormal segment. Multislice CT with multiplanar reformation is the most accurate modality for detecting spondylolysis in adults. Symptoms may occur in the setting of a single injury, repeated minor injuries, or during a growth spurt. Spondylolysis is the most common cause of persistent LBP in adolescents and is often associated with sports-related activities.

Scoliosis refers to an abnormal curvature in the coronal (lateral) plane of the spine. With *kyphoscoliosis,* there is, in addition, a forward curvature of the spine. The abnormal curvature may be congenital, due to abnormal spine development, acquired in adulthood due to degenerative spine disease, or progressive due to paraspinal neuromuscular disease. The deformity can progress until ambulation or pulmonary function is compromised.

Spina bifida occulta (closed spinal dysraphism) is a failure of closure of one or several vertebral arches posteriorly; the meninges and spinal cord are normal. A dimple or small lipoma may overlie the defect, but the skin is intact. Most cases are asymptomatic and discovered incidentally during a physical examination for back pain.

Tethered cord syndrome usually presents as a progressive cauda equina disorder (see below), although myelopathy may also be the initial manifestation. The patient is often a child or young adult who complains of perineal or perianal pain, sometimes following minor trauma. MRI studies typically reveal a low-lying conus (below L1 and L2) and a short and thickened filum terminale. The MRI findings also occur as incidental findings, sometimes during evaluation of unrelated LBP in adults.

REFERRED PAIN FROM VISCERAL DISEASE

Diseases of the thorax, abdomen, or pelvis may refer pain to the spinal segment that innervates the diseased organ. Occasionally, back pain may be the first and only manifestation. Upper abdominal diseases generally refer pain to the lower thoracic or upper lumbar region (eighth thoracic to the first and second lumbar vertebrae), lower abdominal diseases to the midlumbar region (second to fourth lumbar vertebrae), and pelvic diseases to the sacral region. Local signs (pain with spine palpation, paraspinal muscle spasm) are absent, and little or no pain accompanies routine movements.

Low Thoracic or Lumbar Pain with Abdominal Disease

Tumors of the posterior wall of the stomach or duodenum typically produce epigastric pain **(Chaps. 80 and 324)**, but back pain may occur if retroperitoneal extension is present. Fatty foods occasionally induce back pain associated with biliary or pancreatic disease. Pathology in retroperitoneal structures (hemorrhage, tumors, and pyelonephritis) can produce paraspinal pain that radiates to the lower abdomen, groin, or anterior thighs. A mass in the iliopsoas region can produce unilateral lumbar pain with radiation toward the groin, labia, or testicle. The sudden appearance of lumbar pain in a patient receiving anticoagulants should prompt consideration of retroperitoneal hemorrhage.

Isolated LBP occurs in some patients with a contained rupture of an AAA. The classic clinical triad of abdominal pain, shock, and back pain occurs in <20% of patients. The diagnosis may be missed because the symptoms and signs can be nonspecific. Misdiagnoses include nonspecific back pain, diverticulitis, renal colic, sepsis, and myocardial infarction. A careful abdominal examination revealing a pulsatile mass (present in 50–75% of patients)

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Pelvic organs rarely cause isolated LBP. Uterine malposition (retroversion, descensus, and prolapse) may cause traction on the uterosacral ligament.

Access Provided by:

anterior thighs. A mass in the iliopsoas region can produce unilateral lumbar pain with radiation toward the groin, labia, or testicle. The sudden

<20% of patients. The diagnosis may be missed because the symptoms and signs can be nonspecific. Misdiagnoses include nonspecific back pain, diverticulitis, renal colic, sepsis, and myocardial infarction. A careful abdominal examination revealing a pulsatile mass (present in 50–75% of patients) is an important physical finding. Patients with suspected AAA should be evaluated with abdominal ultrasound, CT, or MRI **(Chap. 280)**.

Sacral Pain with Gynecologic and Urologic Disease

Pelvic organs rarely cause isolated LBP. Uterine malposition (retroversion, descensus, and prolapse) may cause traction on the uterosacral ligament. The pain is referred to the sacral region, sometimes appearing after prolonged standing. Endometriosis or uterine cancers can invade the uterosacral ligaments. Pain associated with endometriosis is typically premenstrual and often continues until it merges with menstrual pain.

Menstrual pain with poorly localized, cramping pain can radiate down the legs. LBP that radiates into one or both thighs is common in the last weeks of pregnancy. Continuous and worsening pain unrelieved by rest or at night may be due to neoplastic infiltration of nerves or nerve roots.

Urologic sources of lumbosacral back pain include chronic prostatitis, prostate cancer with spinal metastasis **(Chap. 87)**, and diseases of the kidney or ureter. Infectious, inflammatory, or neoplastic renal diseases may produce ipsilateral lumbosacral pain, as can renal artery or vein thrombosis. Paraspinal lumbar pain may be a symptom of ureteral obstruction due to nephrolithiasis.

OTHER CAUSES OF BACK PAIN

Postural Back Pain

There is a group of patients with nonspecific chronic low back pain (CLBP) in whom no specific anatomic lesion can be found despite exhaustive investigation. Exercises to strengthen the paraspinal and abdominal muscles are sometimes helpful. CLBP may be encountered in patients who seek financial compensation; in malingerers; or in those with concurrent substance abuse. Many patients with CLBP have a history of psychiatric illness (depression, anxiety states) or childhood trauma (physical or sexual abuse) that antedates the onset of back pain. Preoperative psychological assessment has been used to exclude patients with marked psychological impairments that predict a poor surgical outcome from spine surgery.

Idiopathic

The cause of LBP occasionally remains unclear. Some patients have had multiple operations for disk disease. The original indications for surgery may have been questionable, with back pain only, no definite neurologic signs, or a minor disk bulge noted on CT or MRI. Scoring systems based on neurologic signs, psychological factors, physiologic studies, and imaging studies have been devised to minimize the likelihood of unsuccessful surgery.

GLOBAL CONSIDERATIONS

While many of the history and examination features described in this chapter apply to all patients, information regarding the global epidemiology and prevalence of LBP is limited. The Global Burden of Diseases Study 2019 reported that LBP represented the #1 cause overall for total years lived with disability (YLD), and #9 overall as a cause of disability-related life years (DALYs). These numbers increased substantially from 1990 estimates, and with the aging of the population worldwide, the numbers of individuals suffering from LBP are expected to increase further in the future. Although rankings for LBP generally were higher in developed regions, a high burden exists in every part of the world. An area of uncertainty is the degree to which regional differences exist in terms of the specific etiologies of LBP and how these are managed. For example, the most common cause of arachnoiditis in developing countries is a prior spinal infection, but in developed countries the most frequent cause is multiple lumbar spine surgeries.

TREATMENT

Back Pain

Management is considered separately for acute and chronic low back pain syndromes without radiculopathy, and for back pain with radiculopathy.

ACUTE LOW BACK PAIN WITHOUT RADICULOPATHY

This is defined as pain of <12 weeks duration. Full recovery can be expected in >85% of adults with ALBP without leg pain. Most have purely “mechanical” symptoms (i.e., pain that is aggravated by motion and relieved by rest).

The initial assessment is focused on excluding serious causes of spine pathology that require urgent intervention, including infection, cancer, or

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The prognosis of ALBP is generally excellent; however, episodes tend to recur, and as many as two-thirds of patients will experience a second episode within 1 year. Most patients do not seek medical care and improve on their own. Even among those seen in primary care, two-thirds report

Access Provided by:

The initial assessment is focused on excluding serious causes of spine pathology that require urgent intervention, including infection, cancer, or trauma. Risk factors for a serious cause of ALBP are shown in Table 17-1. Laboratory and imaging studies are unnecessary if risk factors are absent. CT, MRI, or plain spine films are rarely indicated in the first month of symptoms unless a spine fracture, tumor, or infection is suspected.

The prognosis of ALBP is generally excellent; however, episodes tend to recur, and as many as two-thirds of patients will experience a second episode within 1 year. Most patients do not seek medical care and improve on their own. Even among those seen in primary care, two-thirds report substantial improvement after 7 weeks. This high likelihood of spontaneous improvement can mislead clinicians and patients about the efficacy of treatment interventions, highlighting the importance of rigorous prospective trials. Many treatments commonly used in the past are now known to be ineffective, including bed rest and lumbar traction.

Clinicians should reassure and educate patients that improvement is very likely and instruct them in self-care. Satisfaction and the likelihood of follow-up increase when patients are educated about prognosis, evidence-based treatments, appropriate activity modifications, and strategies to prevent future exacerbations. Counseling patients about the risks of overtreatment is another important part of the discussion. Patients who report that they did not receive an adequate explanation for their symptoms are likely to request further diagnostic tests.

In general, bed rest should be avoided for relief of severe symptoms or limited to a day or two at most. Several randomized trials suggest that bed rest does not hasten the pace of recovery. In general, early resumption of normal daily physical activity should be encouraged, avoiding only strenuous manual labor. Advantages of early ambulation for ALBP also include maintenance of cardiovascular conditioning; improved bone, cartilage, and muscle strength; and increased endorphin levels. Specific back exercises or early vigorous exercise have not shown benefits for acute back pain. Empiric use of heating pads or blankets is sometimes helpful.

NSAIDs and Acetaminophen

Evidence-based guidelines recommend over-the-counter medicines such as NSAIDs and acetaminophen as first-line options for treatment of ALBP. In otherwise healthy patients, a trial of NSAIDs can be followed by acetaminophen for time-limited periods. In theory, the anti-inflammatory effects of NSAIDs might provide an advantage over acetaminophen to suppress inflammation that accompanies many causes of ALBP, but in practice there is no clinical evidence to support the superiority of NSAIDs. The risk of renal and gastrointestinal toxicity with NSAIDs is increased in patients with preexisting medical comorbidities (e.g., renal insufficiency, cirrhosis, prior gastrointestinal hemorrhage, use of anticoagulants or glucocorticoids, heart failure). Some patients elect to take acetaminophen and an NSAID together in hopes of a more rapid benefit.

Muscle Relaxants

Skeletal muscle relaxants, such as cyclobenzaprine or methocarbamol, may be useful, but sedation is a common side effect. Limiting the use of muscle relaxants to nighttime only may be an option for patients with back pain that interferes with sleep.

Opioids

There is no good evidence to support the use of opioid analgesics or tramadol as first-line therapy for ALBP. Their use is best reserved for patients who cannot tolerate acetaminophen or NSAIDs and for those with severe refractory pain. Also, the duration of opioid treatment for ALBP should be strictly limited to 3–7 days. As with muscle relaxants, these drugs are often sedating, so it may be useful to prescribe them at nighttime only. Side effects of short-term opioid use include nausea, constipation, and pruritus; risks of long-term opioid use include hypersensitivity to pain, hypogonadism, and dependency. Falls, fractures, driving accidents, and fecal impaction are other risks. The clinical efficacy of opioids for chronic pain beyond 16 weeks of use is unproven.

Mounting evidence of morbidity from long-term opioid therapy (including overdose, dependency, addiction, falls, fractures, accident risk, and sexual dysfunction) has prompted efforts to reduce its use for chronic pain, including back pain **(Chap. 13)**. When used, safety may be improved with automated notices for high doses, early refills, prescriptions from multiple pharmacies, overlapping opioid and benzodiazepine prescriptions, and in the United States by state-based prescription drug monitoring programs (PDMPs). A recent study indicated that most patients with opioid use disorder presenting to emergency departments had no prescriptions recorded in the PDMP, reflecting other methods used to obtain opioids. Greater access to alternative treatments for chronic pain, such as tailored exercise programs and cognitive behavioral therapy (CBT), may also reduce opioid prescribing.

Other Approaches

There is no evidence to support use of oral or injected glucocorticoids, antiepileptics, antidepressants, or therapies for neuropathic pain such as gabapentin or herbal therapies. Commonly used nonpharmacologic treatments for ALBP are also of unproven benefit, including spinal

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

symptomatic relief from ice or frozen gel packs just before sleep, and heat may produce a short-term reduction in pain after the first week. Patients often report improved satisfaction with the care that they receive when they actively participate in the selection of symptomatic approaches.

Access Provided by:

There is no evidence to support use of oral or injected glucocorticoids, antiepileptics, antidepressants, or therapies for neuropathic pain such as gabapentin or herbal therapies. Commonly used nonpharmacologic treatments for ALBP are also of unproven benefit, including spinal manipulation, physical therapy, massage, acupuncture, laser therapy, therapeutic ultrasound, corsets, transcutaneous electrical nerve stimulation (TENS), special mattresses, or lumbar traction. Although important for chronic pain, use of back exercises for ALBP are generally not supported by clinical evidence. There is no convincing evidence regarding the value of ice or heat applications for ABLP; however, many patients report temporary symptomatic relief from ice or frozen gel packs just before sleep, and heat may produce a short-term reduction in pain after the first week. Patients often report improved satisfaction with the care that they receive when they actively participate in the selection of symptomatic approaches.

CHRONIC LOW BACK PAIN WITHOUT RADICULOPATHY

Back pain is considered chronic when the symptoms last >12 weeks; it accounts for 50% of total back pain costs. Risk factors include obesity, female gender, older age, prior history of back pain, restricted spinal mobility, pain radiating into a leg, high levels of psychological distress, poor self-rated health, minimal physical activity, smoking, job dissatisfaction, and widespread pain. In general, the same treatments that are recommended for ALBP can be useful for patients with CLBP. In this setting, however, the benefit of opioid therapy or muscle relaxants is less clear. In general, improved activity tolerance is the primary goal, while pain relief is secondary.

Some observers have raised concerns that CLBP may often be overtreated. For CLBP without radiculopathy, multiple guidelines explicitly recommend against use of SSRIs, any type of injection, TENS, lumbar supports, traction, radiofrequency facet joint denervation, intradiskal electrothermal therapy, or intradiskal radiofrequency thermocoagulation. On the other hand, exercise therapy and treatment of depression appear to be useful and underused.

Exercise Programs

Evidence supports the use of exercise therapy to alleviate pain symptoms and improve function. Exercise can be one of the mainstays of treatment for CLBP. Effective regimens have generally included a combination of core-strengthening exercises, stretching, and gradually increasing aerobic exercise. A program of supervised exercise can improve compliance. Supervised intensive physical exercise or “work hardening” regimens have been effective in returning some patients to work, improving walking distance, and reducing pain. In addition, some forms of yoga have been evaluated in randomized trials and may be helpful for patients who are interested.

Intensive multidisciplinary rehabilitation programs can include daily or frequent physical therapy, exercise, CBT, a workplace evaluation, and other interventions. For patients who have not responded to other approaches, such programs appear to offer some benefit. Systematic reviews, however, suggest that the evidence and benefits are limited.

Nonopioid Medications

Medications for CLBP may include short courses of NSAIDs or acetaminophen. Duloxetine is approved for the treatment of CLBP (60 mg daily) and may also treat coincident depression. Tricyclic antidepressants can provide modest pain relief for some patients without evidence of depression. Depression is common among patients with chronic pain and should be appropriately treated.

Cognitive Behavioral Therapy

CBT is based on evidence that psychological and social factors, as well as somatic pathology, are important in the genesis of chronic pain and disability; CBT focuses on efforts to identify and modify patients’ thinking about their condition. In one randomized trial, CBT reduced disability and pain in patients with CLBP. Such behavioral treatments appear to provide benefits similar in magnitude to exercise therapy.

Complementary Medicine

Back pain is the most frequent reason for seeking complementary and alternative treatments. Spinal manipulation or massage therapy may provide short-term relief, but long-term benefit is unproven. Biofeedback has not been studied rigorously. There is no convincing evidence that either TENS, laser therapy, or ultrasound are effective in treating CLBP. Rigorous trials of acupuncture suggest that true acupuncture is not superior to sham acupuncture, but that both may offer an advantage over routine care. Whether this is due entirely to placebo effects provided even by sham acupuncture is uncertain.

Injections and Other Interventions

Various injections, including epidural glucocorticoid injections, facet joint injections, and trigger point injections, have been used for treating CLBP. However, in the absence of radiculopathy, there is no clear evidence that these approaches are sustainably effective.

Injection studies are sometimes used diagnostically to help determine the anatomic source of back pain. Pain relief following a glucocorticoid and

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Another category of intervention for CLBP is electrothermal and radiofrequency therapy. Intradiskal therapy has been proposed using energy to thermocoagulate and destroy nerves in the intervertebral disk, using specially designed catheters or electrodes. Current evidence does not support

Access Provided by:

Injection studies are sometimes used diagnostically to help determine the anatomic source of back pain. Pain relief following a glucocorticoid and anesthetic injection into a facet or medial branch block are used as evidence that the facet joint is the pain source; however, the possibility that the response was a placebo effect or due to systemic absorption of the glucocorticoids is difficult to exclude.

Another category of intervention for CLBP is electrothermal and radiofrequency therapy. Intradiskal therapy has been proposed using energy to thermocoagulate and destroy nerves in the intervertebral disk, using specially designed catheters or electrodes. Current evidence does not support the use of discography to identify a specific disk as the pain source, or the use of intradiskal electrothermal or radiofrequency therapy for CLBP.

Radiofrequency denervation is sometimes used to destroy nerves that are thought to mediate pain, and this technique has been used for facet joint pain (with the target nerve being the medial branch of the primary dorsal ramus), for back pain thought to arise from the intervertebral disk (ramus communicans), and radicular back pain (dorsal root ganglia). These interventional therapies have not been studied in sufficient detail to draw firm conclusions regarding their value for CLBP.

Surgery

Surgical intervention for CLBP without radiculopathy has been evaluated in a number of randomized trials. The case for fusion surgery for CLBP without radiculopathy is weak. While some studies have shown modest benefit, there has been no benefit when compared to an active medical treatment arm, often including highly structured, rigorous rehabilitation combined with CBT. The use of bone matrix protein (BMP) instead of iliac crest graft for the fusion was shown to increase hospital costs and length of stay but not improve clinical outcomes.

Guidelines suggest that referral for an opinion on spinal fusion can be considered for patients who have completed an optimal nonsurgical treatment program (including combined physical and psychological treatment) and who have persistent severe back pain for which they would consider surgery. The high cost, wide geographic variations, and rapidly increasing rates of spinal fusion surgery have prompted scrutiny regarding the lack of standardization of appropriate indications. Some insurance carriers have begun to limit coverage for the most controversial indications, such as LBP without radiculopathy.

Lumbar disk replacement with prosthetic disks is US Food and Drug Administration–approved for uncomplicated patients needing single-level surgery at the L3-S1 levels. The disks are generally designed as metal plates with a polyethylene cushion sandwiched in between. The trials that led to approval of these devices were not blinded. When compared to spinal fusion, the artificial disks were “not inferior.” Long-term follow-up is needed to determine device failure rates over time. Serious complications are somewhat more likely with the artificial disk. This treatment remains controversial for CLBP.

LOW BACK PAIN WITH RADICULOPATHY

A common cause of back pain with radiculopathy is a herniated disk affecting the nerve root and producing back pain with radiation down the leg. The term *sciatica* is used when the leg pain radiates posteriorly in a sciatic or L5/S1 distribution. The prognosis for acute low back and leg pain with radiculopathy due to disk herniation is generally favorable, with most patients showing substantial improvement over months. Serial imaging studies suggest spontaneous regression of the herniated portion of the disk in two-thirds of patients over 6 months. Nonetheless, several important treatment options provide symptomatic relief while the healing process unfolds.

Resumption of normal activity is recommended. Randomized trial evidence suggests that bed rest is ineffective for treating sciatica as well as back pain alone. Acetaminophen and NSAIDs are useful for pain relief, although severe pain may require short courses (3–7 days) of opioid analgesics. Opioids are superior for acute pain relief in the emergency department.

Epidural glucocorticoid injections have a role in providing symptom relief for acute lumbar radiculopathy due to a herniated disk, but do not reduce the use of subsequent surgical intervention. A brief course of high-dose oral glucocorticoids (methylprednisolone dose pack) for 3 days followed by a rapid taper over 4 more days can be helpful for some patients with acute disk-related radiculopathy, although this specific regimen has not been studied rigorously.

Diagnostic nerve root blocks have been advocated to determine if pain originates from a specific nerve root. However, improvement may result even when the nerve root is not responsible for the pain; this may occur as a placebo effect, from a pain-generating lesion located distally along the peripheral nerve, or from effects of systemic absorption.

Urgent surgery is recommended for patients who have evidence of CES or spinal cord compression, generally manifesting as combinations of bowel or bladder dysfunction, diminished sensation in a saddle distribution, a sensory level on the trunk, and bilateral leg weakness or spasticity. Surgical intervention is also indicated for patients with progressive motor weakness due to nerve root injury demonstrated on clinical examination or EMG.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

patient has failed to respond to a minimum of 6–8 weeks of nonsurgical management. For patients who have not improved, randomized trials show that surgery results in more rapid pain relief than nonsurgical treatment. However, after 2 years of follow-up, patients appear to have similar pain

Access Provided by:

or bladder dysfunction, diminished sensation in a saddle distribution, a sensory level on the trunk, and bilateral leg weakness or spasticity. Surgical intervention is also indicated for patients with progressive motor weakness due to nerve root injury demonstrated on clinical examination or EMG.

Surgery is also an important option for patients who have disabling radicular pain despite optimal conservative treatment. Because patients with a herniated disk and sciatica generally experience rapid improvement over weeks, most experts do not recommend considering surgery unless the patient has failed to respond to a minimum of 6–8 weeks of nonsurgical management. For patients who have not improved, randomized trials show that surgery results in more rapid pain relief than nonsurgical treatment. However, after 2 years of follow-up, patients appear to have similar pain relief and functional improvement with or without surgery. Thus, both treatment approaches are reasonable, and patient preferences and needs (e.g., rapid return to employment) strongly influence decision-making. Some patients will want the fastest possible relief and find surgical risks acceptable. Others will be more risk-averse and more tolerant of symptoms and will choose watchful waiting, especially if they understand that improvement is likely in the end.

The usual surgical procedure is a partial hemilaminectomy with excision of the prolapsed disk (diskectomy). Minimally invasive techniques have gained in popularity in recent years, but some evidence suggests they may be less effective than standard surgical techniques, with more residual back pain, leg pain, and higher rates of rehospitalization. Fusion of the involved lumbar segments should be considered only if significant spinal instability is present (i.e., degenerative spondylolisthesis). The costs associated with lumbar interbody fusion have increased dramatically in recent years. There are no large prospective, randomized trials comparing fusion to other types of surgical intervention. In one study, patients with persistent LBP despite an initial diskectomy fared no better with spine fusion than with a conservative regimen of cognitive intervention and exercise. Artificial disks, as discussed above, are used in Europe; their utility remains controversial in the United States.

PAIN IN THE NECK AND SHOULDER

Neck pain, which usually arises from diseases of the cervical spine and soft tissues of the neck, is common, typically precipitated by movement, and may be accompanied by focal tenderness and limitation of motion. Many of the earlier comments made regarding causes of LBP also apply to disorders of the cervical spine. The text below will emphasize differences. Pain arising from the brachial plexus, shoulder, or peripheral nerves can be confused with cervical spine disease **(Table 17-4)**, but the history and examination usually identify a more distal origin for the pain. When the site of nerve tissue injury is unclear, EMG studies can localize the lesion. Cervical spine trauma, disk disease, or spondylosis with intervertebral foraminal narrowing may be asymptomatic or painful and can produce a myelopathy, radiculopathy, or both. The same risk factors for serious causes of LBP also apply to neck pain with the additional feature that neurologic signs of myelopathy (incontinence, sensory level, spastic legs) may also occur. Lhermitte's sign, an electrical shock down the spine with neck flexion, suggests involvement of the cervical spinal cord.

TABLE 17-4

Cervical Radiculopathy: Neurologic Features

CERVICAL NERVE ROOT	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
C5	Biceps	Lateral deltoid	Rhomboids[a] (elbow extends backward with hand on hip)	Lateral arm, medial scapula
			Infraspinatus[a] (arm rotates externally with elbow flexed at the side)	
			Deltoid[a] (arm raised laterally 30°–45° from the side)	
C6	Biceps	Palmar thumb/index finger	Biceps[a] (arm flexed at the elbow in supination)	Lateral forearm, thumb/index fingers
				dorsal hand

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:

Enthesitis (sign), an electrical shock down the spine with neck flexion, suggests involvement of the cervical spinal cord.

TABLE 17-4

Cervical Radiculopathy: Neurologic Features

CERVICAL NERVE ROOT	EXAMINATION FINDINGS			PAIN DISTRIBUTION
	REFLEX	SENSORY	MOTOR	
C5	Biceps	Lateral deltoid	Rhomboids[a] (elbow extends backward with hand on hip)	Lateral arm, medial scapula
			Infraspinatus[a] (arm rotates externally with elbow flexed at the side)	
			Deltoid[a] (arm raised laterally 30°–45° from the side)	
C6	Biceps	Palmar thumb/index finger	Biceps[a] (arm flexed at the elbow in supination)	Lateral forearm, thumb/index fingers
		Dorsal hand/lateral forearm	Pronator teres (forearm pronated)	
C7	Triceps	Middle finger	Triceps[a] (forearm extension, flexed at elbow)	Posterior arm, dorsal forearm, dorsal hand
		Dorsal forearm	Wrist/finger extensors[a]	
C8	Finger flexors	Palmar surface of little finger	Abductor pollicis brevis (abduction of thumb)	Fourth and fifth fingers, medial hand and forearm
		Medial hand and forearm	First dorsal interosseous (abduction of index finger)	
			Abductor digiti minimi (abduction of little finger)	
T1	Finger flexors	Axilla, medial arm, anteromedial forearm	Abductor pollicis brevis (abduction of thumb)	Medial arm, axilla
			First dorsal interosseous (abduction of index finger)	
			Abductor digiti minimi (abduction of little finger)	

[a]These muscles receive the majority of innervation from this root.

TRAUMA TO THE CERVICAL SPINE

Trauma (fractures, subluxation) places the spinal cord at risk for compression. Motor vehicle accidents, violent crimes, or falls account for 87% of cervical spinal cord injuries **(Chap. 442)**. Immediate immobilization of the neck is essential to minimize further spinal cord injury from movement of unstable cervical spine segments. A CT scan is the diagnostic procedure of choice for detection of acute fractures following severe trauma; plain x-rays

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The decision to obtain imaging should be based on the clinical context of the injury. The National Emergency X-Radiography Utilization Study (NEXUS) low-risk criteria established that normally alert patients without palpation tenderness in the midline; intoxication; neurologic deficits; or painful

Access Provided by:

cervical spinal cord injuries **(Chap. 442)**. Immediate immobilization of the neck is essential to minimize further spinal cord injury from movement of unstable cervical spine segments. A CT scan is the diagnostic procedure of choice for detection of acute fractures following severe trauma; plain x-rays are used for lesser degrees of trauma or in settings where CT is unavailable. When traumatic injury to the vertebral arteries or cervical spinal cord is suspected, visualization by MRI with magnetic resonance angiography is preferred.

The decision to obtain imaging should be based on the clinical context of the injury. The National Emergency X-Radiography Utilization Study (NEXUS) low-risk criteria established that normally alert patients without palpation tenderness in the midline; intoxication; neurologic deficits; or painful distracting injuries were very unlikely to have sustained a clinically significant traumatic injury to the cervical spine. The Canadian C-spine rule recommends that imaging should be obtained following neck region trauma if the patient is >65 years old or has limb paresthesias or if there was a dangerous mechanism for the injury (e.g., bicycle collision with tree or parked car, fall from height >3 ft or five stairs, diving accident). These guidelines are helpful but must be tailored to individual circumstances; for example, patients with advanced osteoporosis, glucocorticoid use, or cancer may warrant imaging after even mild trauma.

Whiplash injury is due to rapid flexion and extension of the neck, usually from automobile accidents. The likely mechanism involves injury to the facet joints. This diagnosis should not be applied to patients with fractures, disk herniation, head injury, focal neurologic findings, or altered consciousness. Up to 50% of persons reporting whiplash injury acutely have persistent neck pain 1 year later. When personal compensation for pain and suffering was removed from the Australian health care system, the prognosis for recovery at 1 year improved. Imaging of the cervical spine is not cost-effective acutely but is useful to detect disk herniations when symptoms persist for >6 weeks following the injury. Severe initial symptoms have been associated with a poor long-term outcome.

CERVICAL DISK DISEASE

Degenerative cervical disk disease is very common and usually asymptomatic. Herniation of a lower cervical disk is a common cause of pain or tingling in the neck, shoulder, arm, or hand. Neck pain, stiffness, and a range of motion limited by pain are the usual manifestations. Herniated cervical disks are responsible for ~25% of cervical radiculopathies. Extension and lateral rotation of the neck narrow the ipsilateral intervertebral foramen and may reproduce radicular symptoms (Spurling's sign). In young adults, acute nerve root compression from a ruptured cervical disk is often due to trauma. Cervical disk herniations are usually posterolateral near the lateral recess. Typical patterns of reflex, sensory, and motor changes that accompany cervical nerve root lesions are summarized in Table 17-4. Although the classic patterns are clinically helpful, there are numerous exceptions because (1) there is overlap in sensory function between adjacent nerve roots, (2) symptoms and signs may be evident in only part of the injured nerve root territory, and (3) the location of pain is the most variable of the clinical features.

CERVICAL SPONDYLOSIS

Osteoarthritis of the cervical spine may produce neck pain that radiates into the back of the head, shoulders, or arms, or may be the source of headaches in the posterior occipital region (supplied by the C2-C4 nerve roots). Osteophytes, disk protrusions, or hypertrophic facet or uncovertebral joints may alone or in combination compress one or several nerve roots at the intervertebral foramina; these causes together account for 75% of cervical radiculopathies. The roots most commonly affected are C7 and C6. Narrowing of the spinal canal by osteophytes, ossification of the posterior longitudinal ligament (OPLL), or a large central disk may compress the cervical spinal cord and produce signs of myelopathy alone or radiculopathy with myelopathy (myeloradiculopathy). When little or no neck pain accompanies cervical cord involvement, other diagnoses to be considered include amyotrophic lateral sclerosis **(Chap. 437)**, multiple sclerosis **(Chap. 444)**, spinal cord tumors, or syringomyelia **(Chap. 442)**. Cervical spondylotic myelopathy should be considered even when the patient presents with symptoms or spinal cord signs in the legs only. MRI is the study of choice to define soft tissues in the cervical region including the spinal cord, whereas plain CT is optimal to identify bone pathology including foraminal, lateral recess, OPLL, or spinal canal stenosis. In spondylotic myelopathy, focal enhancement by MRI, sometimes in a characteristic "pancake pattern," may be present at the site of maximal cord compression.

There is no evidence to support prophylactic surgery for asymptomatic cervical spinal stenosis unaccompanied by myelopathic signs or abnormal spinal cord findings on MRI, except in the setting of *dynamic instability* (see spondylolisthesis above). If the patient has postural neck pain, a prior history of whiplash or other spine/head injury, a Lhermitte sign, or preexisting listhesis at the stenotic segment on cervical MRI or CT, then cervical spine flexion-extension x-rays or MRI are indicated to look for dynamic instability. Surgical intervention is not recommended for patients with listhesis alone, unaccompanied by dynamic instability.

OTHER CAUSES OF NECK PAIN

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

routinely performed preoperatively to assess the risk of neck hyperextension in patients requiring intubation. The degree of subluxation correlates with the severity of erosive disease. When subluxation is present, careful assessment is important to identify early signs of myelopathy that could be a

Access Provided by:

OTHER CAUSES OF NECK PAIN

Rheumatoid arthritis (RA) **(Chap. 358)** of the cervical facet joints produces neck pain, stiffness, and limitation of motion. Synovitis of the atlantoaxial joint (C1-C2; Fig. 17-2) may damage the transverse ligament of the atlas, producing forward displacement of the atlas on the axis (atlantoaxial subluxation). Radiologic evidence of atlantoaxial subluxation occurs in up to 30% of patients with RA and plain x-ray films of the neck should be routinely performed preoperatively to assess the risk of neck hyperextension in patients requiring intubation. The degree of subluxation correlates with the severity of erosive disease. When subluxation is present, careful assessment is important to identify early signs of myelopathy that could be a harbinger of life-threatening spinal cord compression. Surgery should be considered when myelopathy or spinal instability is present. *Ankylosing spondylitis* is another cause of neck pain and less commonly atlantoaxial subluxation.

Acute *herpes zoster* can present as acute posterior occipital or neck pain prior to the outbreak of vesicles. *Neoplasms* metastatic to the cervical spine, *infections* (osteomyelitis and epidural abscess), and *metabolic bone diseases* may be the cause of neck pain, as discussed above. Neck pain may also be referred from the heart with coronary artery ischemia (cervical angina syndrome). Rheumatologic disease should be considered if the neck pain is accompanied by shoulder or hip girdle pain.

THORACIC OUTLET SYNDROMES

The thoracic outlet contains the first rib, the subclavian artery and vein, the brachial plexus, the clavicle, and the lung apex. Injury to these structures may result in postural or movement-induced pain around the shoulder and supraclavicular region, classified as follows.

True neurogenic thoracic outlet syndrome (*TOS*) is an uncommon disorder resulting from compression of the lower trunk of the brachial plexus or ventral rami of the C8 or T1 nerve roots, caused most often by an anomalous band of cartilaginous tissue connecting an elongate transverse process at C7 with the first rib. Pain is mild or may be absent. Signs include weakness and wasting of intrinsic muscles of the hand and diminished sensation on the palmar aspect of the fifth digit. An anteroposterior cervical spine x-ray will show an elongate C7 transverse process (an anatomic marker for the anomalous cartilaginous band), and EMG and NCSs confirm the diagnosis. Treatment consists of surgical resection of the anomalous band. The weakness and wasting of intrinsic hand muscles typically do not improve, but surgery halts the insidious progression of weakness.

Arterial TOS results from compression of the subclavian artery by a cervical rib, resulting in poststenotic dilatation of the artery and in some cases secondary thrombus formation. Blood pressure is reduced in the affected limb, and signs of emboli may be present in the hand. Neurologic signs are absent. Ultrasound can confirm the diagnosis noninvasively. Treatment is with thrombolysis or anticoagulation (with or without embolectomy) and surgical excision of the cervical rib compressing the subclavian artery.

Venous TOS is due to subclavian vein thrombosis resulting in swelling of the arm and pain. The vein may be compressed by a cervical rib or anomalous scalene muscle. Venography is the diagnostic test of choice.

Disputed TOS accounts for 95% of patients diagnosed with TOS; chronic arm and shoulder pain are prominent and of unclear cause. The lack of sensitive and specific findings on physical examination or specific markers for this condition results in diagnostic uncertainty. The role of surgery in disputed TOS is controversial. Major depression, chronic symptoms, work-related injury, and diffuse arm symptoms predict poor surgical outcomes. Multidisciplinary pain management is a conservative approach, although treatment is often unsuccessful.

BRACHIAL PLEXUS AND NERVES

Pain from injury to the brachial plexus or peripheral nerves of the arm can occasionally mimic referred pain of cervical spine origin, including cervical radiculopathy, but the pain typically begins distal to the posterior neck region in the shoulder girdle or upper arm. Neoplastic infiltration of the lower trunk of the brachial plexus may produce shoulder or supraclavicular pain radiating down the arm, numbness of the fourth and fifth fingers or medial forearm, and weakness of intrinsic hand muscles innervated by the lower trunk and medial cord of the brachial plexus. Delayed radiation injury may produce weakness in the upper arm or numbness of the lateral forearm or arm due to involvement of the upper trunk and lateral cord of the plexus. Pain is less common and less severe than with neoplastic infiltration. A Pancoast tumor of the lung **(Chap. 78)** is another cause and should be considered, especially when a concurrent Horner's syndrome is present. *Acute brachial neuritis* is often confused with radiculopathy; the acute onset of severe shoulder or scapular pain is followed typically over days by weakness of the proximal arm and shoulder girdle muscles innervated by the upper brachial plexus. The onset may be preceded by an infection, vaccination, or minor surgical procedure. The long thoracic nerve may be affected, resulting in a winged scapula. Brachial neuritis may also present as an isolated paralysis of the diaphragm with or without involvement of other nerves of the upper limb. Recovery may take up to 3 years, and full functional recovery can be expected in the majority of patients.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

For further discussion of peripheral nerve disorders, see Chap. 446.

Access Provided by:

of the upper limb. Recovery may take up to 3 years, and full functional recovery can be expected in the majority of patients.

Occasional cases of carpal tunnel syndrome produce pain and paresthesias extending into the forearm, arm, and shoulder resembling a C5 or C6 root lesion. Lesions of the radial or ulnar nerve can also mimic radiculopathy, at C7 or C8, respectively. EMG and NCSs can accurately localize lesions to the nerve roots, brachial plexus, or peripheral nerves.

For further discussion of peripheral nerve disorders, see Chap. 446.

SHOULDER

Pain arising from the shoulder can on occasion mimic pain from the spine. If symptoms and signs of radiculopathy are absent, then the differential diagnosis includes mechanical shoulder pain (bicipital tendonitis, frozen shoulder, bursitis, rotator cuff tear, dislocation, adhesive capsulitis, or rotator cuff impingement under the acromion) and referred pain (subdiaphragmatic irritation, angina, Pancoast tumor). Mechanical pain is often worse at night, associated with local shoulder tenderness and aggravated by passive abduction, internal rotation, or extension of the arm. Demonstrating normal passive full range of motion of the arm at the shoulder without worsening the usual pain can help exclude mechanical shoulder pathology as a cause of neck region pain. Pain from shoulder disease may radiate into the arm or hand, but focal neurologic signs (sensory, motor, or reflex changes) are absent.

GLOBAL CONSIDERATIONS

Many of the considerations described above for LBP also apply to neck pain. The Global Burden of Diseases Study 2019 reported that neck pain ranked second only to back pain as a cause of total years lived with disability (YLD). In general, neck pain rankings were also higher in developed regions of the world.

TREATMENT

Neck Pain Without Radiculopathy

The evidence regarding treatment for neck pain is less comprehensive than that for LBP, but the approach is remarkably similar in many respects. As with LBP, spontaneous improvement is the norm for acute neck pain. The usual goals of therapy are to promote a rapid return to normal function and provide pain relief while healing proceeds.

Acute neck pain is often treated with NSAIDs, acetaminophen, cold packs, or heat, alone or in combination while awaiting recovery. Patients should be specifically educated regarding the favorable natural history of acute neck pain to avoid unrealistic fear and inappropriate requests for imaging and other tests. For patients kept awake by symptoms, cyclobenzaprine (5–10 mg) at night can help relieve muscle spasm and promote drowsiness. For patients with neck pain unassociated with trauma, supervised exercise with or without mobilization appears to be effective. Exercises often include shoulder rolls and neck stretches. The evidence in support of nonsurgical treatments for whiplash-associated disorders is generally of limited quality and neither supports nor refutes the common treatments used for symptom relief. Gentle mobilization of the cervical spine combined with exercise programs may be beneficial. Evidence is insufficient to recommend the use of cervical traction, TENS, ultrasound, trigger point injections, botulinum toxin injections, tricyclic antidepressants, and SSRIs for acute or chronic neck pain. Some patients obtain modest pain relief using a soft neck collar; there is little risk or cost. Massage can produce temporary pain relief.

For patients with chronic neck pain, supervised exercise programs can provide symptom relief and improve function. Acupuncture provided short-term benefit for some patients when compared to a sham procedure and is an option. Spinal manipulation alone has not been shown to be effective and carries a risk for injury. Surgical treatment for chronic neck pain without radiculopathy or spine instability is not recommended.

Neck Pain With Radiculopathy

The natural history of acute neck pain with radiculopathy due to disk disease is favorable, and many patients will improve without specific therapy. Although there are no randomized trials of NSAIDs for neck pain, a course of NSAIDs, acetaminophen, or both, with or without muscle relaxants, and avoidance of activities that trigger symptoms are reasonable as initial therapy. Gentle supervised exercise and avoidance of inactivity are reasonable as well. A short course of high-dose oral glucocorticoids with a rapid taper, or epidural steroids administered under imaging guidance can be effective for acute or subacute disk-related cervical radicular pain, but have not been subjected to rigorous trials. The risk of injection-related complications is higher in the neck than the low back; vertebral artery dissection, dural puncture, spinal cord injury, and embolism in the vertebral arteries have all been reported. Opioid analgesics can be used in the emergency department and for short courses as an outpatient. Soft cervical

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

If cervical radiculopathy is due to bony compression from cervical spondylosis with foraminal narrowing, periodic follow-up to assess for progression is indicated and consideration of surgical decompression is reasonable. Surgical treatment can produce rapid pain relief, although it is

Access Provided by:

TREATMENT

Neck Pain Without Radiculopathy

The evidence regarding treatment for neck pain is less comprehensive than that for LBP, but the approach is remarkably similar in many respects. As with LBP, spontaneous improvement is the norm for acute neck pain. The usual goals of therapy are to promote a rapid return to normal function and provide pain relief while healing proceeds.

Acute neck pain is often treated with NSAIDs, acetaminophen, cold packs, or heat, alone or in combination while awaiting recovery. Patients should be specifically educated regarding the favorable natural history of acute neck pain to avoid unrealistic fear and inappropriate requests for imaging and other tests. For patients kept awake by symptoms, cyclobenzaprine (5–10 mg) at night can help relieve muscle spasm and promote drowsiness. For patients with neck pain unassociated with trauma, supervised exercise with or without mobilization appears to be effective. Exercises often include shoulder rolls and neck stretches. The evidence in support of nonsurgical treatments for whiplash-associated disorders is generally of limited quality and neither supports nor refutes the common treatments used for symptom relief. Gentle mobilization of the cervical spine combined with exercise programs may be beneficial. Evidence is insufficient to recommend the use of cervical traction, TENS, ultrasound, trigger point injections, botulinum toxin injections, tricyclic antidepressants, and SSRIs for acute or chronic neck pain. Some patients obtain modest pain relief using a soft neck collar; there is little risk or cost. Massage can produce temporary pain relief.

For patients with chronic neck pain, supervised exercise programs can provide symptom relief and improve function. Acupuncture provided short-term benefit for some patients when compared to a sham procedure and is an option. Spinal manipulation alone has not been shown to be effective and carries a risk for injury. Surgical treatment for chronic neck pain without radiculopathy or spine instability is not recommended.

Neck Pain With Radiculopathy

The natural history of acute neck pain with radiculopathy due to disk disease is favorable, and many patients will improve without specific therapy. Although there are no randomized trials of NSAIDs for neck pain, a course of NSAIDs, acetaminophen, or both, with or without muscle relaxants, and avoidance of activities that trigger symptoms are reasonable as initial therapy. Gentle supervised exercise and avoidance of inactivity are reasonable as well. A short course of high-dose oral glucocorticoids with a rapid taper, or epidural steroids administered under imaging guidance can be effective for acute or subacute disk-related cervical radicular pain, but have not been subjected to rigorous trials. The risk of injection-related complications is higher in the neck than the low back; vertebral artery dissection, dural puncture, spinal cord injury, and embolism in the vertebral arteries have all been reported. Opioid analgesics can be used in the emergency department and for short courses as an outpatient. Soft cervical collars can be modestly helpful by limiting spontaneous and reflex neck movements that exacerbate pain; hard collars are in general poorly tolerated.

If cervical radiculopathy is due to bony compression from cervical spondylosis with foraminal narrowing, periodic follow-up to assess for progression is indicated and consideration of surgical decompression is reasonable. Surgical treatment can produce rapid pain relief, although it is unclear if long-term functional outcomes are improved over nonsurgical therapy. Indications for cervical disk surgery include a progressive motor deficit due to nerve root compression, functionally limiting pain that fails to respond to conservative management, or spinal cord compression. In other circumstances, clinical improvement over time regardless of therapeutic intervention is common.

Surgical treatments include anterior cervical diskectomy alone, laminectomy with diskectomy, or diskectomy with fusion. The risk of subsequent radiculopathy or myelopathy at cervical segments adjacent to a fusion is ~3% per year and 26% per decade. Although this risk is sometimes portrayed as a late complication of surgery, it may also reflect the natural history of degenerative cervical disk disease.

FURTHER READING

Agency for Healthcare Research and Quality (AHRQ): Noninvasive treatments for low back pain. AHRQ Publication No. 16-EHC004-EF. February 2016, *https://effectivehealthcare.ahrq.gov/ehc/products/553/2178/back-pain-treatment-report-160229.pdf*

Austevoll IM et al: Decompression with or without fusion in degenerative lumbar spondylolisthesis. N Engl J Med 385:526, 2021. [PubMed: 34347953]

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Burden of Disease Study 2019. Lancet 396:2006, 2021. [PubMed: 33275908]

Access Provided by:

Bailey CS et al: Surgery versus conservative care for persistent sciatica lasting 4 to 12 months. N Engl J Med 19;382:1093, 2020.

Cieza A et al: Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: A systematic analysis for the Global Burden of Disease Study 2019. Lancet 396:2006, 2021. [PubMed: 33275908]

Engstrom JW: Physical and Neurologic Examination. In Steinmetz et al (eds). *Benzel's Spine Surgery*, 5th ed. Philadelphia, Elsevier, 2021.

Goldberg H et al: Oral steroids for acute radiculopathy due to a herniated lumbar disk. JAMA 313:1915, 2015. [PubMed: 25988461]

Hawk K et al: Past-year prescription drug monitoring program opioid prescriptions and self-reported opioid use in an emergency department population with opioid use disorder. Acad Emerg Med 25:508, 2018. [PubMed: 29165853]

Jarvik JG et al: Association of early imaging for back pain with clinical outcomes in older adults. JAMA 313:1143, 2015. [PubMed: 25781443]

Katz JN, Harris MB: Clinical practice. Lumbar spinal stenosis. N Engl J Med 358:818, 2008. [PubMed: 18287604]

Theodore N: Degenerative cervical spondylosis. N Engl J Med 383:159, 2020. [PubMed: 32640134]

Zygourakis CC et al: Geographic and hospital variation in cost of lumbar laminectomy and lumbar fusion for degenerative conditions. Neurosurgery 81:331, 2017. [PubMed: 28327960]

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 17: Back and Neck Pain, John W. Engstrom
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:

Harrison's Principles of Internal Medicine 21e

Chapter 18: Fever

Neeraj K. Surana; Charles A. Dinarello; Reuven Porat

INTRODUCTION

Body temperature is controlled by the hypothalamus. Neurons in both the preoptic anterior hypothalamus and the posterior hypothalamus receive two kinds of signals: one from peripheral nerves that transmit information from warmth/cold receptors in the skin and the other from the temperature of the blood bathing the region. These two types of signals are integrated by the thermoregulatory center of the hypothalamus to maintain normal temperature. In a neutral temperature environment, the human metabolic rate produces more heat than is necessary to maintain the core body temperature in the range of 36.5–37.5°C (97.7–99.5°F).

A normal body temperature is ordinarily maintained despite environmental variations because the hypothalamic thermoregulatory center balances the excess heat production derived from metabolic activity in muscle and the liver with heat dissipation from the skin and lungs. According to a study of >35,000 individuals ≥18 years of age seen in routine medical visits, the mean oral temperature is 36.6°C (95% confidence interval, 35.7–37.3°C). In light of this study, *a temperature of >37.7°C (>99.9°F), which represents the 99th percentile for healthy individuals, defines a fever.* Importantly, higher ambient temperatures are linked to higher baseline body temperatures. Additionally, body temperatures have diurnal and seasonal variation, with low levels at 8 A.M. and during summer and higher levels at 4 P.M. and during winter. Baseline temperatures are also affected by age (lower by 0.02°C for every 10-year increase in age), demographics (African-American women have temperatures 0.052°C higher than white men), and comorbid conditions (cancer is associated with 0.02°C higher temperatures; hypothyroidism is linked to temperatures lower by 0.01°C). After controlling for age, sex, race, vital signs, and comorbidities, an increase in baseline temperature of 0.15°C (or 1 standard deviation) intriguingly translates into a 0.52% absolute increase in 1-year mortality.

Rectal temperatures are generally 0.4°C (0.7°F) higher than oral readings. The lower oral readings are probably attributable to mouth breathing, which is a factor in patients with respiratory infections and rapid breathing. Lower-esophageal temperatures closely reflect core temperature. Tympanic membrane thermometers measure radiant heat from the tympanic membrane and nearby ear canal and display that absolute value (*unadjusted mode*) or a value automatically calculated from the absolute reading on the basis of nomograms relating the radiant temperature measured to actual core temperatures obtained in clinical studies (*adjusted mode*). These measurements, although convenient, may be more variable than directly determined oral or rectal values. Studies in adults show that readings are lower with unadjusted-mode than with adjusted-mode tympanic membrane thermometers and that unadjusted-mode tympanic membrane values are 0.8°C (1.6°F) lower than rectal temperatures.

In women who menstruate, the A.M. temperature is generally lower during the 2 weeks before ovulation; it then rises by ~0.6°C (1°F) with ovulation and stays at that level until menses occur. During the luteal phase, the amplitude of the circadian rhythm remains the same.

FEVER VERSUS HYPERTHERMIA

Fever is an elevation of body temperature that exceeds the normal daily variation and occurs *in conjunction with an increase in the hypothalamic set point* (e.g., from 37°C to 39°C). This shift of the set point from "normothermic" to febrile levels very much resembles the resetting of the home thermostat to a higher level in order to raise the ambient temperature in a room. Once the hypothalamic set point is raised, neurons in the vasomotor center are activated and vasoconstriction commences. The individual first notices vasoconstriction in the hands and feet. Shunting of blood away from the periphery to the internal organs essentially decreases heat loss from the skin, and the person feels cold. For most fevers, body temperature increases by 1–2°C. Shivering, which increases heat production from the muscles, may begin at this time; however, shivering is not required if mechanisms of heat conservation raise blood temperature sufficiently. Nonshivering heat production from the liver also contributes to increasing core temperature. Behavioral adjustments (e.g., putting on more clothing or bedding) help raise body temperature by decreasing heat loss.

The processes of heat conservation (vasoconstriction) and heat production (shivering and increased nonshivering thermogenesis) continue until the

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

through vasodilation and sweating are initiated. Loss of heat by sweating and vasodilation continues until the blood temperature at the hypothalamic level matches the lower setting. Behavioral changes (e.g., removal of clothing) facilitate heat loss.

Access Provided by:

The processes of heat conservation (vasoconstriction) and heat production (shivering and increased nonshivering thermogenesis) continue until the temperature of the blood bathing the hypothalamic neurons matches the new "thermostat setting." Once that point is reached, the hypothalamus maintains the temperature at the febrile level by the same mechanisms of heat balance that function in the afebrile state. When the hypothalamic set point is again reset downward (in response to either a reduction in the concentration of pyrogens or the use of antipyretics), the processes of heat loss through vasodilation and sweating are initiated. Loss of heat by sweating and vasodilation continues until the blood temperature at the hypothalamic level matches the lower setting. Behavioral changes (e.g., removal of clothing) facilitate heat loss.

A fever of >41.5°C (>106.7°F) is called *hyperpyrexia*. This extraordinarily high fever can develop in patients with severe infections but most commonly occurs in patients with central nervous system (CNS) hemorrhages. In the preantibiotic era, fever due to a variety of infectious diseases rarely exceeded 106°F, and there has been speculation that this natural "thermal ceiling" is mediated by neuropeptides functioning as central antipyretics.

In rare cases, the hypothalamic set point is elevated as a result of local trauma, hemorrhage, tumor, or intrinsic hypothalamic malfunction. The term *hypothalamic fever* is sometimes used to describe elevated temperature caused by abnormal hypothalamic function. However, most patients with hypothalamic damage have *sub*normal, not *supra*normal, body temperatures.

Although most patients with elevated body temperature have fever, there are circumstances in which elevated temperature represents not fever but *hyperthermia* (*heat stroke*). Hyperthermia is characterized by an uncontrolled increase in body temperature that exceeds the body's ability to lose heat. The setting of the hypothalamic thermoregulatory center is unchanged. In contrast to fever in infections, hyperthermia does not involve pyrogenic molecules. Exogenous heat exposure and endogenous heat production are two mechanisms by which hyperthermia can result in dangerously high internal temperatures. Excessive heat production can easily cause hyperthermia despite physiologic and behavioral control of body temperature. For example, work or exercise in hot environments can produce heat faster than peripheral mechanisms can lose it. **For a detailed discussion of hyperthermia, see Chap. 465.**

It is important to distinguish between fever and hyperthermia since hyperthermia can be rapidly fatal and characteristically does not respond to antipyretics. In an emergency situation, however, making this distinction can be difficult. For example, in systemic sepsis, fever (hyperpyrexia) can be rapid in onset, and temperatures can exceed 40.5°C (104.9°F). Hyperthermia is often diagnosed on the basis of the events immediately preceding the elevation of core temperature—e.g., heat exposure or treatment with drugs that interfere with thermoregulation. In patients with heat stroke syndromes and in those taking drugs that block sweating, the skin is hot but dry, whereas in fever, the skin can be cold as a consequence of vasoconstriction. Antipyretics do not reduce the elevated temperature in hyperthermia, whereas in fever—and even in hyperpyrexia—adequate doses of either aspirin or acetaminophen usually result in some decrease in body temperature.

PATHOGENESIS OF FEVER

PYROGENS

The term *pyrogen* (Greek *pyro*, "fire") is used to describe any substance that causes fever. *Exogenous* pyrogens are derived from outside the patient; most are microbial products, microbial toxins, or whole microorganisms (including viruses). The classic example of an exogenous pyrogen is the lipopolysaccharide (endotoxin) produced by all gram-negative bacteria. Pyrogenic products of gram-positive organisms include the enterotoxins of *Staphylococcus aureus* and the groups A and B streptococcal toxins, also called *superantigens.* One staphylococcal toxin of clinical importance is that associated with isolates of *S. aureus* from patients with toxic shock syndrome. These products of staphylococci and streptococci cause fever in experimental animals when injected intravenously at concentrations of 1–10 μg/kg. Endotoxin is a highly pyrogenic molecule in humans: when injected intravenously into volunteers, a dose of 2–3 ng/kg produces fever, leukocytosis, acute-phase proteins, and generalized symptoms of malaise.

PYROGENIC CYTOKINES

Cytokines are small proteins (molecular mass, 10,000–20,000 Da) that regulate immune, inflammatory, and hematopoietic processes. For example, the elevated leukocytosis seen in several infections with an absolute neutrophilia is attributable to the cytokines interleukin (IL) 1 and IL-6. Some cytokines also cause fever; formerly referred to as *endogenous pyrogens*, they are now called *pyrogenic cytokines.* The pyrogenic cytokines include IL-1, IL-6, tumor necrosis factor (TNF), and ciliary neurotropic factor, a member of the IL-6 family. Fever is a prominent side effect of interferon α therapy. Each pyrogenic cytokine is encoded by a separate gene, and each has been shown to cause fever in laboratory animals and in humans. When injected into humans at low doses (10–100 ng/kg), IL-1 and TNF produce fever; in contrast, for IL-6, a dose of 1–10 μg/kg is required for fever production.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 18: Fever, Neeraj K. Surana; Charles A. Dinarello; Reuven Porat
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

levels.

Access Provided by:

humans at low doses (10–100 ng/kg), IL-1 and TNF produce fever; in contrast, for IL-6, a dose of 1–10 μg/kg is required for fever production.

A wide spectrum of bacterial and fungal products induce the synthesis and release of pyrogenic cytokines. However, fever can be a manifestation of disease in the absence of microbial infection. For example, inflammatory processes such as pericarditis, trauma, stroke, and routine immunizations induce the production of IL-1, TNF, and/or IL-6; individually or in combination, these cytokines trigger the hypothalamus to raise the set point to febrile levels.

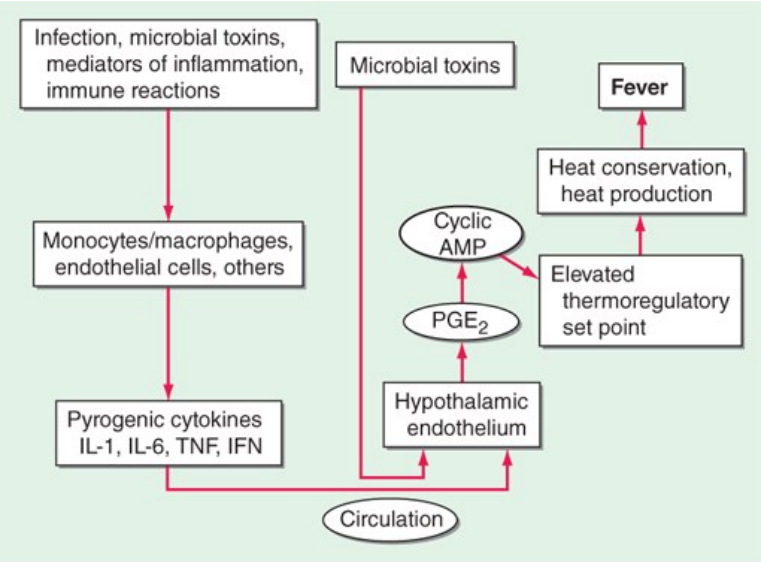
ELEVATION OF THE HYPOTHALAMIC SET POINT BY CYTOKINES

During fever, levels of prostaglandin E_2 (PGE_2) are elevated in hypothalamic tissue and the third cerebral ventricle. The concentrations of PGE_2 are highest near the circumventricular vascular organs (organum vasculosum of lamina terminalis)—networks of enlarged capillaries surrounding the hypothalamic regulatory centers. Destruction of these organs reduces the ability of pyrogens to produce fever. Most studies in animals have failed to show, however, that pyrogenic cytokines pass from the circulation into the brain itself. Thus, it appears that both exogenous pyrogens and pyrogenic cytokines interact with the endothelium of these capillaries and that this interaction is the first step in initiating fever—i.e., in raising the set point to febrile levels.

The key events in the production of fever are illustrated in **Fig. 18-1**. Myeloid and endothelial cells are the primary cell types that produce pyrogenic cytokines. Pyrogenic cytokines such as IL-1, IL-6, and TNF are released from these cells and enter the systemic circulation. Although these circulating cytokines lead to fever by inducing the synthesis of PGE_2, they also induce PGE_2 in peripheral tissues. The increase in PGE_2 in the periphery accounts for the nonspecific myalgias and arthralgias that often accompany fever. It is thought that some systemic PGE_2 escapes destruction by the lung and gains access to the hypothalamus via the internal carotid. However, it is the elevation of PGE_2 in the brain that starts the process of raising the hypothalamic set point for core temperature.

FIGURE 18-1

Chronology of events required for the induction of fever. AMP, adenosine 5′-monophosphate; IFN, interferon; IL, interleukin; PGE_2, prostaglandin E_2; TNF, tumor necrosis factor.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

There are four receptors for PGE_2, and each signals the cell in different ways. Of the four receptors, the third (EP-3) is essential for fever: when the gene for this receptor is deleted in mice, no fever follows the injection of IL-1 or endotoxin. Deletion of the other PGE_2 receptor genes leaves the fever mechanism intact. Although PGE_2 is essential for fever, it is not a neurotransmitter. Rather, the release of PGE_2 from the brain side of the hypothalamic endothelium triggers the PGE_2 receptor on glial cells, and this stimulation results in the rapid release of cyclic adenosine 5′-monophosphate (cAMP), which is a neurotransmitter. As shown in Fig. 18-1, the release of cAMP from glial cells activates neuronal endings from the thermoregulatory center

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 18: Fever, Neeraj K. Surana; Charles A. Dinarello; Reuven Porat
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

mechanism. Thus, the direct activation of Toll-like receptors or IL-1 receptors results in PGE_2 production and fever.

There are four receptors for PGE_2, and each signals the cell in different ways. Of the four receptors, the third (EP-3) is essential for fever: when the gene

Access Provided by:

endothelium triggers the PGE_2 receptor on glial cells, and this stimulation results in the rapid release of cyclic adenosine 5′-monophosphate (cAMP), which is a neurotransmitter. As shown in Fig. 18-1, the release of cAMP from glial cells activates neuronal endings from the thermoregulatory center that extend into the area. The elevation of cAMP is thought to account for changes in the hypothalamic set point either directly or indirectly (by inducing the release of neurotransmitters). Distinct receptors for microbial products are located on the hypothalamic endothelium. These receptors are called *Toll-like receptors* and are similar in many ways to IL-1 receptors. IL-1 receptors and Toll-like receptors share the same signal-transducing mechanism. Thus, the direct activation of Toll-like receptors or IL-1 receptors results in PGE_2 production and fever.

PRODUCTION OF CYTOKINES IN THE CNS

Cytokines produced in the brain may account for the hyperpyrexia of CNS hemorrhage, trauma, or infection. Viral infections of the CNS induce microglial and possibly neuronal production of IL-1, TNF, and IL-6. In experimental animals, the concentration of a cytokine required to cause fever is several orders of magnitude lower with direct injection into the brain substance or brain ventricles than with systemic injection. Therefore, cytokines produced in the CNS can raise the hypothalamic set point, bypassing the circumventricular organs. CNS cytokines likely account for the hyperpyrexia of CNS hemorrhage, trauma, or infection.

APPROACH TO THE PATIENT

Fever

HISTORY AND PHYSICAL EXAMINATION

There are a range of disease processes that present with fever as a cardinal manifestation, and a thorough history can help distinguish between these broad categories **(Table 18-1)**. The chronology of events preceding fever, including exposure to other symptomatic individuals or to vectors of disease, should be ascertained. Electronic devices for measuring oral, tympanic membrane, or rectal temperatures are reliable, but the same site should be used consistently to monitor a febrile disease. Moreover, physicians should be aware that newborns, elderly patients, patients with chronic hepatic or renal failure, and patients taking glucocorticoids or being treated with an anticytokine may have active disease in the absence of fever because of a blunted febrile response.

LABORATORY TESTS

The workup should include a complete blood count; a differential count should be performed manually or with an instrument sensitive to the identification of juvenile or band forms, toxic granulations, and Döhle bodies, which are suggestive of bacterial infection. Neutropenia may be present with some viral infections.

Measurement of circulating cytokines in patients with fever is not helpful since levels of cytokines such as IL-1 and TNF in the circulation often are below the detection limit of the assay or do not coincide with fever. However, in patients with low-grade fevers or with suspected occult disease, the most valuable measurements are the C-reactive protein (CRP) level and the erythrocyte sedimentation rate. These markers of inflammatory processes are particularly helpful in detecting occult disease. Measurement of circulating IL-6, which induces CRP, can be useful. However, whereas IL-6 levels may vary during a febrile disease, CRP levels remain elevated. **Acute-phase reactants are discussed in Chap. 304.**

FEVER IN PATIENTS RECEIVING ANTICYTOKINE THERAPY

Patients receiving long-term treatment with anticytokine-based regimens are at increased risk of infection because of lowered host defenses. For example, latent *Mycobacterium tuberculosis* infection can disseminate in patients receiving anti-TNF therapy. With the increasing use of anticytokines to reduce the activity of IL-1, IL-6, IL-12, IL-17, or TNF in patients with Crohn's disease, rheumatoid arthritis, or psoriasis, the possibility that these therapies blunt the febrile response should be kept in mind.

The blocking of cytokine activity has the distinct clinical drawback of lowering the level of host defenses against both routine bacterial and opportunistic infections such as *M. tuberculosis* and fungal infections. The use of monoclonal antibodies to reduce IL-17 in psoriasis increases the risk of systemic candidiasis.

In nearly all reported cases of infection associated with anticytokine therapy, fever is among the presenting signs. However, the extent to which the febrile response is blunted in these patients remains unknown. Therefore, low-grade fever in patients receiving anticytokine therapies is of considerable concern. The physician should conduct an early and rigorous diagnostic evaluation in these cases. The febrile response is also blunted in patients receiving chronic glucocorticoid therapy or anti-inflammatory agents such as nonsteroidal anti-inflammatory drugs (NSAIDs).

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 18: Fever, Neeraj K. Surana; Charles A. Dinarello; Reuven Porat

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Disease Categories That Present with Fever as a Cardinal Sign

Access Provided by:

APPROACH TO THE PATIENT

Fever

HISTORY AND PHYSICAL EXAMINATION

There are a range of disease processes that present with fever as a cardinal manifestation, and a thorough history can help distinguish between these broad categories **(Table 18-1)**. The chronology of events preceding fever, including exposure to other symptomatic individuals or to vectors of disease, should be ascertained. Electronic devices for measuring oral, tympanic membrane, or rectal temperatures are reliable, but the same site should be used consistently to monitor a febrile disease. Moreover, physicians should be aware that newborns, elderly patients, patients with chronic hepatic or renal failure, and patients taking glucocorticoids or being treated with an anticytokine may have active disease in the absence of fever because of a blunted febrile response.

LABORATORY TESTS

The workup should include a complete blood count; a differential count should be performed manually or with an instrument sensitive to the identification of juvenile or band forms, toxic granulations, and Döhle bodies, which are suggestive of bacterial infection. Neutropenia may be present with some viral infections.

Measurement of circulating cytokines in patients with fever is not helpful since levels of cytokines such as IL-1 and TNF in the circulation often are below the detection limit of the assay or do not coincide with fever. However, in patients with low-grade fevers or with suspected occult disease, the most valuable measurements are the C-reactive protein (CRP) level and the erythrocyte sedimentation rate. These markers of inflammatory processes are particularly helpful in detecting occult disease. Measurement of circulating IL-6, which induces CRP, can be useful. However, whereas IL-6 levels may vary during a febrile disease, CRP levels remain elevated. **Acute-phase reactants are discussed in Chap. 304**.

FEVER IN PATIENTS RECEIVING ANTICYTOKINE THERAPY

Patients receiving long-term treatment with anticytokine-based regimens are at increased risk of infection because of lowered host defenses. For example, latent *Mycobacterium tuberculosis* infection can disseminate in patients receiving anti-TNF therapy. With the increasing use of anticytokines to reduce the activity of IL-1, IL-6, IL-12, IL-17, or TNF in patients with Crohn's disease, rheumatoid arthritis, or psoriasis, the possibility that these therapies blunt the febrile response should be kept in mind.

The blocking of cytokine activity has the distinct clinical drawback of lowering the level of host defenses against both routine bacterial and opportunistic infections such as *M. tuberculosis* and fungal infections. The use of monoclonal antibodies to reduce IL-17 in psoriasis increases the risk of systemic candidiasis.

In nearly all reported cases of infection associated with anticytokine therapy, fever is among the presenting signs. However, the extent to which the febrile response is blunted in these patients remains unknown. Therefore, low-grade fever in patients receiving anticytokine therapies is of considerable concern. The physician should conduct an early and rigorous diagnostic evaluation in these cases. The febrile response is also blunted in patients receiving chronic glucocorticoid therapy or anti-inflammatory agents such as nonsteroidal anti-inflammatory drugs (NSAIDs).

TABLE 18-1

Disease Categories That Present with Fever as a Cardinal Sign

Infectious diseases
Autoimmune and noninfectious inflammatory disorders
Cancer
Medication related (e.g., vaccines, drug fever)
Endocrine disorders (e.g., hyperthyroidism)

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 18: Fever, Neeraj K. Surana; Charles A. Dinarello; Reuven Porat
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:

TABLE 18-1

Disease Categories That Present with Fever as a Cardinal Sign

Infectious diseases
Autoimmune and noninfectious inflammatory disorders
Cancer
Medication related (e.g., vaccines, drug fever)
Endocrine disorders (e.g., hyperthyroidism)
Intrinsic hypothalamic malfunction

TREATMENT

Fever

THE DECISION TO TREAT FEVER

In deciding whether to treat fever, it is important to remember that fever itself is not an illness: it is an ordinary response to a perturbation of normal host physiology. Most fevers are associated with self-limited infections, such as common viral diseases. The use of antipyretics is not contraindicated in these infections: no significant clinical evidence indicates either that antipyretics delay the resolution of viral or bacterial infections or that fever facilitates recovery from infection or acts as an adjuvant to the immune system. In short, treatment of fever and its symptoms with routine antipyretics does no harm and does not slow the resolution of common viral and bacterial infections.

However, in bacterial infections, the withholding of antipyretic therapy can be helpful in evaluating the effectiveness of a particular antibiotic, especially in the absence of positive cultures of the infecting organism, and the routine use of antipyretics can mask an inadequately treated bacterial infection. Withholding antipyretics in some cases may facilitate the diagnosis of an unusual febrile disease. Temperature–pulse dissociation (*relative bradycardia*) occurs in typhoid fever, brucellosis, leptospirosis, some drug-induced fevers, and factitious fever. As stated earlier, in newborns, elderly patients, patients with chronic liver or kidney failure, and patients taking glucocorticoids, fever may not be present despite infection. Hypothermia can develop in patients with septic shock.

Some infections have characteristic patterns in which febrile episodes are separated by intervals of normal temperature. For example, *Plasmodium vivax* causes fever every third day, whereas fever occurs every fourth day with *Plasmodium malariae*. Another relapsing fever is related to *Borrelia* infection, with days of fever followed by a several-day afebrile period and then a relapse into additional days of fever. In the Pel-Ebstein pattern, fever lasting 3–10 days is followed by afebrile periods of 3–10 days; this pattern can be classic for Hodgkin's disease and other lymphomas. In cyclic neutropenia, fevers occur every 21 days and accompany the neutropenia. There are also a number of periodic fever syndromes (e.g., familial Mediterranean fever, TNF receptor–associated periodic syndrome [TRAPS]) that differ in their periodicity, duration of attack, constellation of clinical features, genetic causes, and therapies **(Chap. 369)**. Understanding these clinical differences can help tailor diagnostic testing to confirm the diagnosis and guide therapy.

ANTICYTOKINE THERAPY TO REDUCE FEVER IN AUTOIMMUNE AND AUTOINFLAMMATORY DISEASES

Recurrent fever is documented at some point in most autoimmune diseases and many autoinflammatory diseases, which include the periodic fever syndromes as well as disorders of inflammasomes (e.g., NLRP3, pyrin) and other components of the innate immune system **(Chap. 349)**. Although fever can be a manifestation of autoimmune diseases, recurrent fevers are characteristic of autoinflammatory diseases, including uncommon diseases such as adult and juvenile Still's disease, familial Mediterranean fever, and hyper-IgD syndrome but also common diseases such as idiopathic pericarditis and gout. In addition to recurrent fevers, neutrophilia and serosal inflammation characterize autoinflammatory diseases. The fevers associated with many of these illnesses are dramatically reduced by blocking of IL-1 activity with anakinra or canakinumab. Anticytokines

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 18: Fever, Neeraj K. Surana; Charles A. Dinarello; Reuven Porat
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

MECHANISMS OF ANTIPYRETIC AGENTS

The reduction of fever by lowering of the elevated hypothalamic set point is a direct function of reduction of the PGE_2 level in the thermoregulatory

Access Provided by:

idiopathic pericarditis and gout. In addition to recurrent fevers, neutrophilia and serosal inflammation characterize autoinflammatory diseases. The fevers associated with many of these illnesses are dramatically reduced by blocking of IL-1 activity with anakinra or canakinumab. Anticytokines therefore reduce fever in autoimmune and autoinflammatory diseases. Although fevers in autoinflammatory diseases are mediated by IL-1β, patients also respond to antipyretics.

MECHANISMS OF ANTIPYRETIC AGENTS

The reduction of fever by lowering of the elevated hypothalamic set point is a direct function of reduction of the PGE_2 level in the thermoregulatory center. The synthesis of PGE_2 depends on the constitutively expressed enzyme cyclooxygenase. The substrate for cyclooxygenase is arachidonic acid released from the cell membrane, and this release is the rate-limiting step in the synthesis of PGE_2. Therefore, inhibitors of cyclooxygenase are potent antipyretics. The antipyretic potency of various drugs is directly correlated with the inhibition of brain cyclooxygenase. Acetaminophen is a poor cyclooxygenase inhibitor in peripheral tissue and lacks noteworthy anti-inflammatory activity; in the brain, however, acetaminophen is oxidized by the P450 cytochrome system, and the oxidized form inhibits cyclooxygenase activity. Moreover, in the brain, the inhibition of another enzyme, COX-3, by acetaminophen may account for the antipyretic effect of this agent. However, COX-3 is not found outside the CNS.

Oral aspirin and acetaminophen are equally effective in reducing fever in humans. NSAIDs such as ibuprofen and specific inhibitors of COX-2 also are excellent antipyretics. Chronic, high-dose therapy with antipyretics such as aspirin or any NSAID does not reduce normal core body temperature. Thus, PGE_2 appears to play no role in normal thermoregulation.

As effective antipyretics, glucocorticoids act at two levels. First, similar to the cyclooxygenase inhibitors, glucocorticoids reduce PGE_2 synthesis by inhibiting the activity of phospholipase A_2, which is needed to release arachidonic acid from the cell membrane. Second, glucocorticoids block the transcription of the mRNA for the pyrogenic cytokines. Limited experimental evidence indicates that ibuprofen and COX-2 inhibitors reduce IL-1-induced IL-6 production and may contribute to the antipyretic activity of NSAIDs.

REGIMENS FOR THE TREATMENT OF FEVER

The objectives in treating fever are first to reduce the elevated hypothalamic set point and second to facilitate heat loss. Reducing fever with antipyretics also reduces systemic symptoms of headache, myalgias, and arthralgias.

Oral aspirin and NSAIDs effectively reduce fever but can adversely affect platelets and the gastrointestinal tract. Therefore, acetaminophen is preferred as an antipyretic. In children, acetaminophen or oral ibuprofen must be used because aspirin increases the risk of Reye's syndrome. If the patient cannot take oral antipyretics, parenteral preparations of NSAIDs and rectal suppositories of various antipyretics can be used.

Treatment of fever in some patients is highly recommended. Fever increases the demand for oxygen (i.e., for every increase of 1°C over 37°C, there is a 13% increase in oxygen consumption) and can aggravate the condition of patients with preexisting impairment of cardiac, pulmonary, or CNS function. Children with a history of febrile or nonfebrile seizure should be aggressively treated to reduce fever. However, it is unclear what triggers the febrile seizure, and there is no correlation between absolute temperature elevation and onset of a febrile seizure in susceptible children.

In hyperpyrexia, the use of cooling blankets facilitates the reduction of temperature; however, cooling blankets should not be used without oral antipyretics. In hyperpyretic patients with CNS disease or trauma (CNS bleeding), reducing core temperature mitigates the detrimental effects of high temperature on the brain.

For a discussion of treatment for hyperthermia, see Chap. 465.

FURTHER READING

Dinarello CA et al: Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. Nature Rev 11:633, 2012.

Gattorno M et al: Classification criteria for autoinflammatory recurrent fevers. Ann Rheum Dis 78:1025, 2019. [PubMed: 31018962]

Kullenberg T et al: Long-term safety profile of anakinra in patients with severe cryopyrin-associated periodic syndromes. Rheumatology 55:1499, 2016. [PubMed: 27143789]

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 18: Fever, Neeraj K. Surana; Charles A. Dinarello; Reuven Porat
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:

2016. [PubMed: 27143789]

Sakkat A et al: Temperature control in critically ill patients with fever: A meta-analysis of randomized controlled trials. J Crit Care 61:89, 2021. [PubMed: 33157310]

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 18: Fever, Neeraj K. Surana; Charles A. Dinarello; Reuven Porat
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 20: Fever of Unknown Origin

Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer

INTRODUCTION

DEFINITION

Clinicians commonly refer to any febrile illness without an initially obvious etiology as *fever of unknown origin* (FUO). Most febrile illnesses either resolve before a diagnosis can be made or develop distinguishing characteristics that lead to a diagnosis. The term *FUO* should be reserved for prolonged febrile illnesses without an established etiology despite intensive evaluation and diagnostic testing. This chapter focuses on FUO in the adult patient.

FUO was originally defined by Petersdorf and Beeson in 1961 as an illness of >3 weeks' duration with fever of ≥38.3°C (≥101°F) on two occasions and an uncertain diagnosis despite 1 week of inpatient evaluation. Nowadays, most patients with FUO are hospitalized only if their clinical condition requires it, and not for diagnostic purposes alone; thus the in-hospital evaluation requirement has been eliminated from the definition. The definition of FUO has been further modified by the exclusion of immunocompromised patients, whose workup requires an entirely different diagnostic and therapeutic approach. For optimal comparison of patients with FUO in different geographic areas, it has been proposed that the quantitative criterion (diagnosis uncertain after 1 week of evaluation) be changed to a qualitative criterion that requires the performance of a specific list of investigations. Accordingly, FUO is now defined as follows:

1. Fever ≥38.3°C (≥101°F) on at least two occasions
2. Illness duration of ≥3 weeks
3. No known immunocompromised state
4. Diagnosis that remains uncertain after a thorough history-taking, physical examination, and the following obligatory investigations: determination of erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) level; platelet count; leukocyte count and differential; measurement of levels of hemoglobin, electrolytes, creatinine, total protein, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, creatine kinase, ferritin, antinuclear antibodies, and rheumatoid factor; protein electrophoresis; urinalysis; blood cultures ($n = 3$); urine culture; chest x-ray; abdominal ultrasonography; and tuberculin skin test (TST) or interferon γ release assay (IGRA).

Closely related to FUO is *inflammation of unknown origin (IUO)*, which has the same definition as FUO, except for the body temperature criterion: IUO is defined as the presence of elevated inflammatory parameters (CRP or ESR) on multiple occasions for a period of at least 3 weeks in an immunocompetent patient with normal body temperature, for which a final explanation is lacking despite history-taking, physical examination, and the obligatory tests listed above. It has been shown that the causes and workup for IUO are the same as for FUO. Therefore, for convenience, the term FUO will refer to both FUO and IUO within the remainder of this chapter.

ETIOLOGY AND EPIDEMIOLOGY

Table 20-1 summarizes the findings of large studies on FUO conducted over the past 20 years.

TABLE 20-1

Etiology of FUO: Pooled Results of Large Studies Published in the Past 20 Years (1999–2019)

	NO. OF		INFECTIONS,	NONINFECTIOUS	MALIGNANCY,	MISCELLANEOUS,	NO
Western	10	1820	17	25	10	10	37

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 20-1

Etiology of FUO: Pooled Results of Large Studies Published in the Past 20 Years (1999–2019)

GEOGRAPHIC AREA	NO. OF COHORTS (INCLUSION PERIOD)	NO. OF PATIENTS	INFECTIONS, MEDIAN % (RANGE)	NONINFECTIOUS INFLAMMATORY DISEASES, MEDIAN % (RANGE)	MALIGNANCY, MEDIAN % (RANGE)	MISCELLANEOUS, MEDIAN % (RANGE)	NO DIAGNOSIS, MEDIAN % (RANGE)
Western Europe	10 (1990–2014)	1820	17 (11–32)	25 (12–32)	10 (3–20)	10 (0–15)	37 (26–51)
Other European and Turkey	13 (1984–2015)	1316	38 (26–59)	25 (15–38)	14 (5–19)	6 (2–18)	16 (4–35)
Middle East	3 2009–2010 and ?[a]	1235	66 (42–79)	15 (7–17)	7 (1–30)	1 (0–12)	8 (2–12)
Asia	20 (1994–2017)	3802	42 (11–58)	20 (7–57)	13 (6–22)	9 (0–15)	18 (0–36)

[a]One study (published in 2015) did not report the inclusion period.

Abbreviation: NIID, non-infectious inflammatory disease.

For references, see supplementary material at Harrison's Online.

The range of FUO etiologies has evolved since its first definition as a result of changes in the spectrum of diseases causing FUO, the widespread use of antibiotics, and especially the availability of new diagnostic techniques. The proportion of cases caused by intraabdominal abscesses and tumors, for example, has decreased because of earlier detection by CT and ultrasound. In addition, infective endocarditis is a less frequent cause because blood culture and echocardiographic techniques have improved. Conversely, some diagnoses such as acute HIV infection were unknown six decades ago.

Roughly comparable to 60 years ago, in non-Western cohorts infections remain the most common cause of FUO. Up to half of all infections in patients with FUO outside Western nations are caused by *Mycobacterium tuberculosis*, which is a less common cause in Western Europe and probably also in the United States. Recent data from the latter, however, have not been reported. In Western cohorts, noninfectious inflammatory diseases (NIIDs), including autoimmune, autoinflammatory, and granulomatous diseases, as well as vasculitides, form the most common cause of FUO. More than one-third of Western patients with FUO have a diagnosis that falls within the category of NIIDs. The number of FUO patients diagnosed with NIIDs probably will not decrease in the near future, as fever may precede more typical manifestations or laboratory evidence of these diseases by months. Moreover, many NIIDs can be diagnosed only after prolonged observation and exclusion of other diseases.

In Western cohorts, FUO remains unexplained in more than one-third of patients. This is much higher than 60 years ago. This difference can be explained by the fact that in patients with fever a diagnosis is often established before 3 weeks have elapsed because these patients tend to seek medical advice earlier, and because better diagnostic techniques, such as CT, MRI, and positron emission tomography (PET)/CT, are now available. Therefore, only the cases that are most difficult to diagnose continue to meet the criteria for FUO. Furthermore, most patients who have FUO without a diagnosis currently do well. A less aggressive diagnostic approach may be used in clinically stable patients once diseases with immediate therapeutic or prognostic consequences have been ruled out. In patients with recurrent fever (defined as repeated episodes of fever interspersed with fever-free intervals of at least 2 weeks and apparent remission of the underlying disease), the chance of attaining an etiologic diagnosis is <50%.

DIFFERENTIAL DIAGNOSIS

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

disease (*Tropheryma whipplei* infection) are quite rare but should always be kept in mind as a cause of FUO since the presenting symptoms can be nonspecific. Serologic testing for Q fever, which results from exposure to animals or animal products, should be performed by immunofluorescence

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

DIFFERENTIAL DIAGNOSIS

The differential diagnosis for FUO is extensive. It is important to remember that FUO is far more often caused by an atypical presentation of a rather common disease than by a very rare disease. **Table 20-2** presents an overview of possible causes of FUO. Atypical presentations of endocarditis, diverticulitis, vertebral osteomyelitis, and extrapulmonary tuberculosis are the more common infectious disease diagnoses. Q fever and Whipple's disease (*Tropheryma whipplei* infection) are quite rare but should always be kept in mind as a cause of FUO since the presenting symptoms can be nonspecific. Serologic testing for Q fever, which results from exposure to animals or animal products, should be performed by immunofluorescence assay (IFA) when the patient lives in a rural area or has a history of heart valve disease, an aortic aneurysm, or a vascular prosthesis. In patients with unexplained symptoms localized to the central nervous system, gastrointestinal tract, or joints, polymerase chain reaction testing for *Tropheryma whipplei* should be performed. Travel to or (former) residence in tropical countries or the American Southwest should lead to consideration of infectious diseases such as malaria, leishmaniasis, histoplasmosis, or coccidioidomycosis. Fever with signs of endocarditis and negative blood culture results poses a special problem. Culture-negative endocarditis **(Chap. 128)** may be due to difficult-to-culture bacteria such as nutritionally variant bacteria, HACEK organisms (including *Haemophilus parainfluenzae*, *H. paraphrophilus*, *Aggregatibacter actinomycetemcomitans*, *A. aphrophilus*, *A. paraphrophilus*, *Cardiobacterium hominis*, *C. valvarum*, *Eikenella corrodens*, and *Kingella kingae*; discussed below), *Coxiella burnetii*, *T. whipplei*, and *Bartonella* species. Marantic endocarditis is a sterile thrombotic disease that occurs as a paraneoplastic phenomenon, especially with adenocarcinomas. Sterile endocarditis is also seen in the context of systemic lupus erythematosus and antiphospholipid syndrome.

TABLE 20-2

All Reported Causes of Fever of Unknown Origin (FUO)[a]

Infections	
Bacterial, nonspecific	Abdominal abscess, adnexitis, apical granuloma, appendicitis, cholangitis, cholecystitis, diverticulitis, endocarditis, endometritis, epidural abscess, infected joint prosthesis, infected vascular catheter, infected vascular prosthesis, infectious arthritis, infective myonecrosis, intracranial abscess, liver abscess, lung abscess, malakoplakia, mastoiditis, mediastinitis, mycotic aneurysm, osteomyelitis, pelvic inflammatory disease, prostatitis, pyelonephritis, pylephlebitis, renal abscess, septic phlebitis, sinusitis, spondylodiscitis, xanthogranulomatous urinary tract infection
Bacterial, specific	Actinomycosis, atypical mycobacterial infection, bartonellosis, brucellosis, *Campylobacter* infection, *Chlamydia pneumoniae* infection, chronic meningococcemia, ehrlichiosis, gonococcemia, legionellosis, leptospirosis, listeriosis, louse-borne relapsing fever (*Borrelia recurrentis*), Lyme disease, melioidosis (*Pseudomonas pseudomallei*), *Mycoplasma* infection, nocardiosis, psittacosis, Q fever (*Coxiella burnetii),* rickettsiosis, *Spirillum minor* infection, *Streptobacillus moniliformis* infection, syphilis, tick-borne relapsing fever (*Borrelia duttonii*), tuberculosis, tularemia, typhoid fever and other salmonelloses, Whipple's disease (*Tropheryma whipplei*), yersiniosis
Fungal	Aspergillosis, blastomycosis, candidiasis, coccidioidomycosis, cryptococcosis, histoplasmosis, *Malassezia furfur* infection, paracoccidioidomycosis, *Pneumocystis jirovecii* pneumonia, sporotrichosis, zygomycosis
Parasitic	Amebiasis, babesiosis, echinococcosis, fascioliasis, malaria, schistosomiasis, strongyloidiasis, toxocariasis, toxoplasmosis, trichinellosis, trypanosomiasis, visceral leishmaniasis
Viral	Colorado tick fever, coxsackievirus infection, cytomegalovirus infection, dengue, Epstein-Barr virus infection, hantavirus infection, hepatitis (A, B, C, D, E), herpes simplex, HIV infection, human herpesvirus 6 infection, parvovirus infection, West Nile virus infection
Noninfectious Inflammatory Diseases	
Systemic rheumatic and autoimmune diseases	Ankylosing spondylitis, antiphospholipid syndrome, autoimmune hemolytic anemia, autoimmune hepatitis, Behçet's disease, cryoglobulinemia, dermatomyositis, Felty syndrome, gout, mixed connective-tissue disease, polymyositis, pseudogout, reactive arthritis, relapsing polychondritis, rheumatic fever, rheumatoid arthritis, Sjögren's syndrome, systemic lupus erythematosus, Vogt-Koyanagi-Harada syndrome
Vasculitis	Allergic vasculitis, eosinophilic granulomatosis with polyangiitis, giant cell vasculitis/polymyalgia rheumatica, granulomatosis with

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

diseases	

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Systemic	Ankylosing spondylitis, antiphospholipid syndrome, autoimmune hemolytic anemia, autoimmune hepatitis, Behçet's disease,
Vasculitis	Allergic vasculitis, eosinophilic granulomatosis with polyangiitis, giant cell vasculitis/polymyalgia rheumatica, granulomatosis with polyangiitis, hypersensitivity vasculitis, Kawasaki disease, polyarteritis nodosa, Takayasu arteritis, urticarial vasculitis
Granulomatous diseases	Idiopathic granulomatous hepatitis, sarcoidosis
Autoinflammatory syndromes	Adult-onset Still's disease, Blau syndrome, CAPS[b] (cryopyrin-associated periodic syndromes), Crohn's disease, DIRA (deficiency of the interleukin 1 receptor antagonist), familial Mediterranean fever, hemophagocytic syndrome, hyper-IgD syndrome (HIDS, also known as mevalonate kinase deficiency), juvenile idiopathic arthritis, PAPA syndrome (pyogenic sterile arthritis, pyoderma gangrenosum, and acne), PFAPA syndrome (periodic fever, aphthous stomatitis, pharyngitis, adenitis), recurrent idiopathic pericarditis, SAPHO (synovitis, acne, pustulosis, hyperostosis, osteomyelitis), Schnitzler syndrome, TRAPS (tumor necrosis factor receptor–associated periodic syndrome)
Neoplasms	
Hematologic malignancies	Amyloidosis, angioimmunoblastic lymphoma, Castleman's disease, Hodgkin's disease, hypereosinophilic syndrome, leukemia, lymphomatoid granulomatosis, malignant histiocytosis, multiple myeloma, myelodysplastic syndrome, myelofibrosis, non-Hodgkin's lymphoma, plasmacytoma, systemic mastocytosis, vaso-occlusive crisis in sickle cell disease
Solid tumors	Most solid tumors and metastases can cause fever. Those most commonly causing FUO are breast, colon, hepatocellular, lung, pancreatic, and renal cell carcinomas.
Benign tumors	Angiomyolipoma, cavernous hemangioma of the liver, craniopharyngioma, necrosis of dermoid tumor in Gardner's syndrome
Miscellaneous Causes	
	ADEM (acute disseminated encephalomyelitis), adrenal insufficiency, aneurysms, anomalous thoracic duct, aortic dissection, aortic-enteral fistula, aseptic meningitis (Mollaret's syndrome), atrial myxoma, brewer's yeast ingestion, Caroli disease, cholesterol emboli, cirrhosis, complex partial status epilepticus, cyclic neutropenia, drug fever, Erdheim-Chester disease, extrinsic allergic alveolitis, Fabry's disease, factitious disease, fire-eater's lung, fraudulent fever, Gaucher disease, Hamman-Rich syndrome (acute interstitial pneumonia), Hashimoto's encephalopathy, hematoma, hypersensitivity pneumonitis, hypertriglyceridemia, hypothalamic hypopituitarism, idiopathic normal-pressure hydrocephalus, inflammatory pseudotumor, Kikuchi's disease, linear IgA dermatosis, mesenteric fibromatosis, metal fume fever, milk protein allergy, myotonic dystrophy, nonbacterial osteitis, organic dust toxic syndrome, panniculitis, POEMS (polyneuropathy, organomegaly, endocrinopathy, monoclonal protein, skin changes), polymer fume fever, post–cardiac injury syndrome, primary biliary cirrhosis, primary hyperparathyroidism, pulmonary embolism, pyoderma gangrenosum, retroperitoneal fibrosis, Rosai-Dorfman disease, sclerosing mesenteritis, silicone embolization, subacute thyroiditis (de Quervain's), Sweet syndrome (acute febrile neutrophilic dermatosis), thrombosis, tubulointerstitial nephritis and uveitis syndrome (TINU), ulcerative colitis
Thermoregulatory Disorders	
Central	Brain tumor, cerebrovascular accident, encephalitis, hypothalamic dysfunction
Peripheral	Anhidrotic ectodermal dysplasia, exercise-induced hyperthermia, hyperthyroidism, pheochromocytoma

[a]This table includes all causes of FUO that have been described in the literature.

[b]CAPS includes chronic infantile neurologic cutaneous and articular syndrome (CINCA, also known as neonatal-onset multisystem inflammatory disease, or NOMID), familial cold autoinflammatory syndrome (FCAS), and Muckle-Wells syndrome.

Of the NIIDs, adult-onset Still's disease, large-vessel vasculitis, polymyalgia rheumatica, systemic lupus erythematodus (SLE), and sarcoidosis are

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Although most tumors can present with fever, malignant lymphoma is by far the most common diagnosis of FUO among the neoplasms. Sometimes

Of the NIIDs, adult-onset Still's disease, large-vessel vasculitis, polymyalgia rheumatica, systemic lupus erythematodus (SLE), and sarcoidosis are rather common diagnoses in patients with FUO. The hereditary autoinflammatory syndromes are very rare (with the exception of familial Mediterranean fever in specific geographic regions) and usually present in young patients. Schnitzler syndrome, which can present at any age, is uncommon but can often be diagnosed easily in a patient with FUO who presents with urticaria, bone pain, and monoclonal gammopathy.

Although most tumors can present with fever, malignant lymphoma is by far the most common diagnosis of FUO among the neoplasms. Sometimes the fever even precedes lymphadenopathy detectable by physical examination.

Apart from drug-induced fever and exercise-induced hyperthermia, none of the miscellaneous causes of fever is found very frequently in patients with FUO. Virtually all drugs can cause fever, even after long-term use. *Drug-induced fever*, including DRESS (drug reaction with eosinophilia and systemic symptoms; **Fig. A1-48**), is often accompanied by eosinophilia and also by lymphadenopathy, which can be extensive. More common causes of drug-induced fever are allopurinol, carbamazepine, lamotrigine, phenytoin, sulfasalazine, furosemide, antimicrobial drugs (especially sulfonamides, minocycline, vancomycin, β-lactam antibiotics, and isoniazid), some cardiovascular drugs (e.g., quinidine), and some antiretroviral drugs (e.g., nevirapine). *Exercise-induced hyperthermia* **(Chaps. 18 and 465)** is characterized by an elevated body temperature that is associated with moderate to strenuous exercise lasting from half an hour up to several hours without an increase in CRP level or ESR. Unlike patients with fever, these patients typically sweat during the temperature elevation. *Factitious fever* (fever artificially induced by the patient—for example, by IV injection of contaminated water) should be considered in all patients but is more common among young women in health-care professions. In *fraudulent fever*, the patient is normothermic but manipulates the thermometer. Simultaneous measurements at different body sites (rectum, ear, mouth) should rapidly identify this diagnosis. Another clue to fraudulent fever is dissociation between pulse rate and temperature.

Previous studies of FUO have shown that a cause is more likely to be found in elderly patients than in younger age groups. In many cases, FUO in the elderly results from an atypical manifestation of a common disease, among which giant cell arteritis and polymyalgia rheumatica are most frequently involved. Tuberculosis is the most common infectious disease associated with FUO in elderly patients, occurring much more often than in younger patients. As many of these diseases are treatable, it is well worth pursuing the cause of fever in elderly patients.

APPROACH TO THE PATIENT

Fever of Unknown Origin

FIRST-STAGE DIAGNOSTIC TESTS

Figure 20-1 shows a structured approach to patients presenting with FUO. The most important step in the diagnostic workup is the search for potentially diagnostic clues (PDCs) through complete and repeated history-taking and physical examination and the obligatory investigations listed above and in the figure. *PDCs* are defined as all localizing signs, symptoms, and abnormalities potentially pointing toward a diagnosis. Although PDCs are often misleading, only with their help can a concise list of probable diagnoses be made. The history should include information about the fever pattern (continuous or recurrent) and duration, previous medical history, present and recent drug use, family history, sexual history, country of origin, recent and remote travel, unusual environmental exposures associated with travel or hobbies, and animal contacts. A complete physical examination should be performed, with special attention to the eyes, lymph nodes, temporal arteries, liver, spleen, sites of previous surgery, entire skin surface, and mucous membranes. Before further diagnostic tests are initiated, antibiotic and glucocorticoid treatment, which can mask many diseases, should be stopped. For example, blood and other cultures are not reliable when samples are obtained during antibiotic treatment, and the size of enlarged lymph nodes usually decreases during glucocorticoid treatment, regardless of the cause of lymphadenopathy. Despite the high percentage of false-positive ultrasounds and the relatively low sensitivity of chest x-rays, the performance of these simple, low-cost diagnostic tests remains obligatory in all patients with FUO in order to separate cases that are caused by easily diagnosed diseases from those that are not. Abdominal ultrasound is preferred to abdominal CT as an obligatory test because of relatively low cost, lack of radiation burden, and absence of side effects.

Only rarely do biochemical tests (beyond the obligatory tests needed to classify a patient's fever as FUO) lead directly to a definitive diagnosis in the absence of PDCs. The diagnostic yield of immunologic serology other than that included in the obligatory tests is relatively low. These tests more often yield false-positive rather than true-positive results and are of little use without PDCs pointing to specific immunologic disorders. Given the absence of specific symptoms in many patients and the relatively low cost of the test, investigation of cryoglobulins appears to be a valuable screening test in patients with FUO.

Multiple blood samples should be cultured in the laboratory long enough to ensure ample growth time for any fastidious organisms, such as HACEK cultures is useful only when previously cultured samples were collected during antibiotic treatment or within 1 week after its discontinuation. FUO

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Multiple blood samples should be cultured in the laboratory long enough to ensure ample growth time for any fastidious organisms, such as HACEK organisms. It is critical to inform the laboratory of the intent to test for unusual organisms. Specialized media should be used when the history suggests uncommon microorganisms, such as *Histoplasma* or *Legionella*. Performing more than three blood cultures or more than one urine culture is useless in patients with FUO in the absence of PDCs (e.g., a high level of clinical suspicion of endocarditis). Repeating blood or urine cultures is useful only when previously cultured samples were collected during antibiotic treatment or within 1 week after its discontinuation. FUO with headache should prompt microbiologic examination of cerebrospinal fluid (CSF) for organisms including herpes simplex virus (especially type 2), *Cryptococcus neoformans*, and *Mycobacterium tuberculosis*. In central nervous system tuberculosis, the CSF typically has elevated protein and lowered glucose concentrations, with a mononuclear pleocytosis. CSF protein levels range from 100 to 500 mg/dL in most patients, the CSF glucose concentration is <45 mg/dL in 80% of cases, and the usual CSF cell count is between 100 and 500 cells/µL.

Microbiologic serology should not be included in the diagnostic workup of patients without PDCs for specific infections. A tuberculin skin test (TST) or interferon γ release assay (IGRA, QuantiFERON test) is included in the obligatory investigations, but it may yield false-negative results in patients with miliary tuberculosis, malnutrition, or immunosuppression. Although the IGRA is less influenced by prior vaccination with bacille Calmette-Guérin (BCG) or by infection with nontuberculous mycobacteria, its sensitivity is similar to that of the TST; a negative TST or IGRA therefore does not exclude a diagnosis of tuberculosis. Miliary tuberculosis is especially difficult to diagnose. Granulomatous disease in liver or bone marrow biopsy samples, for example, should always lead to a (re)consideration of this diagnosis. If miliary tuberculosis is suspected, liver biopsy for acid-fast smear, culture, and polymerase chain reaction probably still has the highest diagnostic yield; however, biopsies of bone marrow, lymph nodes, or other involved organs also can be considered.

The diagnostic yield of echocardiography, sinus radiography, radiologic or endoscopic evaluation of the gastrointestinal tract, and bronchoscopy is very low in the absence of PDCs. Therefore, these tests should not be used as screening procedures.

After identification of all PDCs retrieved from the history, physical examination, and obligatory tests, a limited list of the most probable diagnoses should be made. Since most investigations are helpful only for patients who have PDCs for the diagnoses sought, further diagnostic procedures should be limited to specific investigations aimed at confirming or excluding diseases on this list. In FUO, the diagnostic pointers are numerous and diverse but may be missed on initial examination, often being detected only by a very careful examination performed subsequently. In the absence of PDCs, the history and physical examination should therefore be repeated regularly. One of the first steps should be to rule out factitious or fraudulent fever, particularly in patients without signs of inflammation in laboratory tests. All medications, including nonprescription drugs and nutritional supplements, should be discontinued early in the evaluation to exclude drug fever. If fever persists beyond 72 h after discontinuation of the suspected drug, it is unlikely that this drug is the cause. In patients without PDCs or with only misleading PDCs, fundoscopy by an ophthalmologist may be useful in the early stage of the diagnostic workup to exclude retinal vasculitis. When the first-stage diagnostic tests do not lead to a diagnosis, ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography combined with computed tomography (PET/CT) or, if the former is not available, radiolabeled leukocyte scintigraphy should be performed, especially when the ESR or the CRP level is elevated.

Recurrent Fever

In patients with recurrent fever, the diagnostic workup should consist of thorough history-taking, physical examination, and obligatory tests. The search for PDCs should be directed toward clues matching known recurrent syndromes **(Table 20-3)**. Patients should be asked to return during a febrile episode so that the history, physical examination, and laboratory tests can be repeated during a symptomatic phase. Further diagnostic tests, such as PET/CT or scintigraphic imaging (see below), should be performed only during a febrile episode or when inflammatory parameters are abnormal because abnormalities may be absent between episodes. In patients with recurrent fever lasting >2 years, it is very unlikely that the fever is caused by infection or malignancy. Further diagnostic tests in that direction should be considered only when PDCs for infections, vasculitis syndromes, or malignancy are present or when the patient's clinical condition is deteriorating.

Fluorodeoxyglucose Positron Emission Tomography

^{18}F-FDG PET/CT has become an established imaging procedure in FUO. FDG accumulates in tissues with a high rate of glycolysis, which occurs not only in malignant cells but also in activated leukocytes and thus permits the imaging of acute and chronic inflammatory processes. Compared with conventional scintigraphy (see below), FDG-PET/CT offers the advantages of higher resolution, greater sensitivity in chronic low-grade infections, and a high degree of accuracy in the central skeleton. Furthermore, vascular uptake of FDG is increased in patients with vasculitis **(Fig. 20-2)**. The mechanisms responsible for FDG uptake do not allow differentiation among infection, sterile inflammation, and malignancy. However, since all of these disorders are causes of FUO, FDG-PET/CT can be used to guide additional diagnostic tests (e.g., targeted biopsies) that may yield the final diagnosis. It is important to realize that physiologic uptake of FDG may obscure pathologic foci in the brain, heart, bowel, kidneys, and bladder. FDG

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

In recent years, many cohort studies and several meta-analyses have focused on the diagnostic yield of PET and PET/CT in FUO. These studies are

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

these disorders are causes of FUO, FDG-PET/CT can be used to guide additional diagnostic tests (e.g., targeted biopsies) that may yield the final diagnosis. It is important to realize that physiologic uptake of FDG may obscure pathologic foci in the brain, heart, bowel, kidneys, and bladder. FDG uptake in the heart, which obscures endocarditis, may be prevented by consumption of a low-carbohydrate diet before the PET investigation. In patients with fever, bone marrow uptake is frequently increased in a nonspecific way due to cytokine activation, which upregulates glucose transporters in bone marrow cells.

In recent years, many cohort studies and several meta-analyses have focused on the diagnostic yield of PET and PET/CT in FUO. These studies are highly variable in terms of the selection of patients, the follow-up, and the selection of a gold-standard reference. Indirect comparisons of test performance suggested that FDG-PET/CT outperformed stand-alone FDG-PET, gallium scintigraphy, and leukocyte scintigraphy. Similarly, indirect comparisons of diagnostic yield suggested that FDG-PET/CT was more likely than alternative tests to correctly identify the cause of FUO. Meta-analyses report a high diagnostic yield for PET and PET/CT in the workup of FUO patients, with pooled sensitivity and specificity figures of ~85% and ~50%, respectively, and a total diagnostic yield of ~50% for PET/CT and ~40% for PET.

As many patients with FUO present with periodic fever, correct timing of PET/CT increases its diagnostic value. Few studies on the use of biomarkers such as elevated CRP or ESR for a contributory outcome of PET/CT have been performed. When both CRP and ESR are normal at the time of FDG-PET/CT, outcome may only be contributory when a patient does have fever at the time of the scan.

Although PET/CT and other scintigraphic techniques do not directly provide a definitive diagnosis (with the exception of some patients with, for instance, large vessel vasculitis), they often identify the anatomic location of a particular ongoing metabolic process. With the help of other techniques such as biopsy and culture, a timely diagnosis and treatment can be facilitated. Pathologic FDG uptake is quickly eradicated by treatment with glucocorticoids in many diseases, including vasculitis and lymphoma; therefore, glucocorticoid use should be stopped or postponed until after FDG-PET/CT is performed.

FDG-PET/CT is a relatively expensive procedure whose availability is still limited compared with that of CT and conventional scintigraphy. Nevertheless, FDG-PET/CT can be cost-effective in the FUO diagnostic workup if used at an early stage, helping to establish an early diagnosis, reducing days of hospitalization for diagnostic purposes, and obviating unnecessary and unhelpful tests. When FDG-PET/CT has been made under the right conditions (i.e., when elevated CRP or ESR or fever were present during the scan) but has not contributed to the final diagnosis, repeating PET/CT is probably of little value, unless new signs or symptoms appear.

Conventional scintigraphic imaging other than PET/CT

Conventional scintigraphic methods used in clinical practice are ^{67}Ga-citrate scintigraphy and ^{111}In- or ^{99m}Tc-labeled leukocyte scintigraphy. Sensitivity and specificity of conventional scintigraphic studies are lower than for PET/CT: the diagnostic yield of gallium scintigraphy ranges from 21% to 54%, and on average the location of a source of fever can correctly be localized in approximately one-third of patients. The diagnostic value of leukocyte scintigraphy ranges from 8% to 31%, and overall the cause of FUO can correctly be identified in one-fifth of patients. When PET/CT is not available, these techniques are the only alternative.

LATER-STAGE DIAGNOSTIC TESTS

In some cases, more invasive tests are appropriate. Abnormalities found with imaging often need to be confirmed by pathology and/or culture of biopsy specimens. If lymphadenopathy is found, lymph node biopsy is necessary, even when the affected lymph nodes are hard to reach or when previous biopsies were inconclusive. In the case of skin lesions, skin biopsy should be undertaken.

If no diagnosis is reached despite PET/CT and PDC-driven histologic investigations or culture, second-stage screening diagnostic tests should be considered (Fig. 20-1). In three studies, the diagnostic yield of screening chest and abdominal CT in patients with FUO was ~20%. The specificity of chest CT was ~80%, but that of abdominal CT varied between 63% and 80%. Despite the relatively limited specificity of abdominal CT and the probably limited additional value of chest CT after normal FDG-PET/CT, chest and abdominal CT may be used as screening procedures at a later stage of the diagnostic protocol because of their noninvasive nature and high sensitivity. Bone marrow aspiration is seldom useful in the absence of PDCs for bone marrow disorders. With addition of FDG-PET/CT, which is highly sensitive in detecting lymphoma, carcinoma, and osteomyelitis, the value of bone marrow biopsy as a screening procedure is probably further reduced. Several studies have shown a high prevalence of giant cell arteritis among patients with FUO, with rates up to 17% among elderly patients. Giant cell arteritis often involves large arteries and in most cases can be diagnosed by FDG-PET/CT. However, temporal artery biopsy is still recommended for patients ≥55 years of age in a later stage of the diagnostic protocol: FDG-PET/CT will not be useful in vasculitis limited to the temporal arteries because of the small diameter of these vessels and the high levels of FDG uptake in the brain. In the past, liver biopsies were often performed as a screening procedure in patients with FUO. In each of two studies, liver biopsy as part of the later stage of a screening diagnostic protocol was helpful in only one patient. Moreover, abnormal liver tests are

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

In patients with unexplained fever after all of the above procedures, the last steps in the diagnostic workup—with only a marginal diagnostic yield—come at an extraordinarily high cost in terms of both expense and discomfort for the patient. Repetition of a thorough history-taking and physical

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

levels of FDG uptake in the brain. In the past, liver biopsies were often performed as a screening procedure in patients with FUO. In each of two studies, liver biopsy as part of the later stage of a screening diagnostic protocol was helpful in only one patient. Moreover, abnormal liver tests are not predictive of a diagnostic liver biopsy in FUO. Liver biopsy is an invasive procedure that carries the possibility of complications and even death. Therefore, it should not be used for screening purposes in patients with FUO except in those with PDCs for liver disease or miliary tuberculosis.

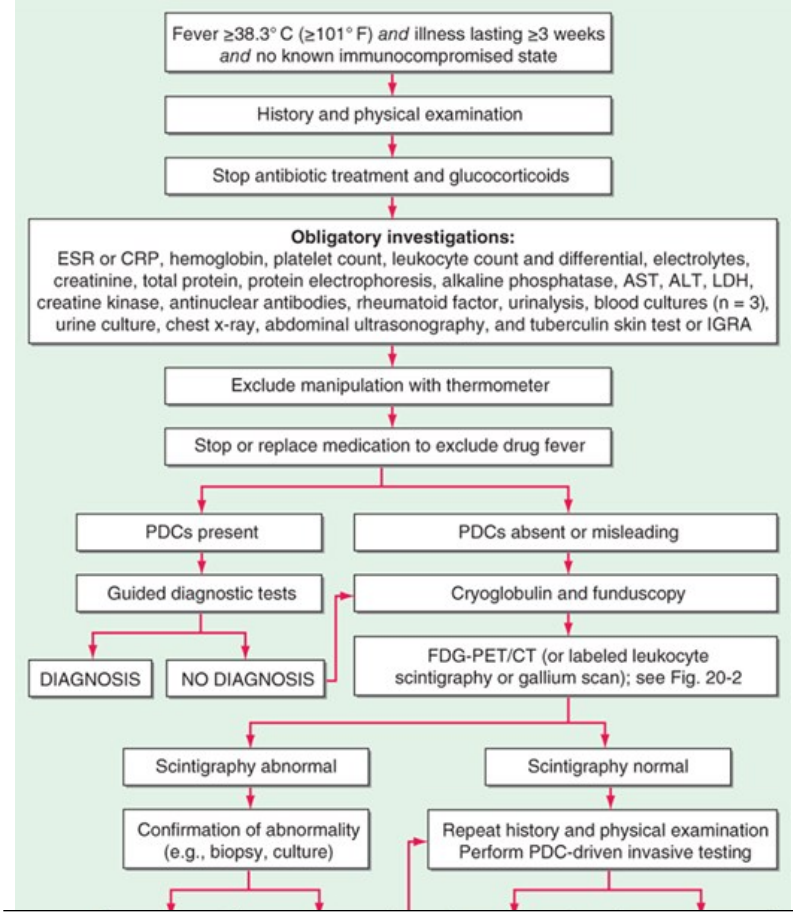
In patients with unexplained fever after all of the above procedures, the last steps in the diagnostic workup—with only a marginal diagnostic yield—come at an extraordinarily high cost in terms of both expense and discomfort for the patient. Repetition of a thorough history-taking and physical examination and review of laboratory results and imaging studies (including those from other hospitals) are recommended. Diagnostic delay often results from a failure to recognize PDCs in the available information. In these patients with persisting FUO, waiting for new PDCs to appear probably is better than ordering more screening investigations. Only when a patient's condition deteriorates without providing new PDCs should a further diagnostic workup be performed.

SECOND OPINION IN AN EXPERT CENTER

When despite the workup described above no explanation for FUO is found, second opinion in an expert center on FUO should be considered. The single study on the value of second opinion in FUO reported that in 57.3% of patients with unexplained FUO, a diagnosis could be found in an expert center. Additionally, of all patients who remained without a diagnosis even after second opinion, 10.9% became fever-free upon empirical treatment, adding up to a beneficial outcome in 68.2% of patients.

FIGURE 20-1

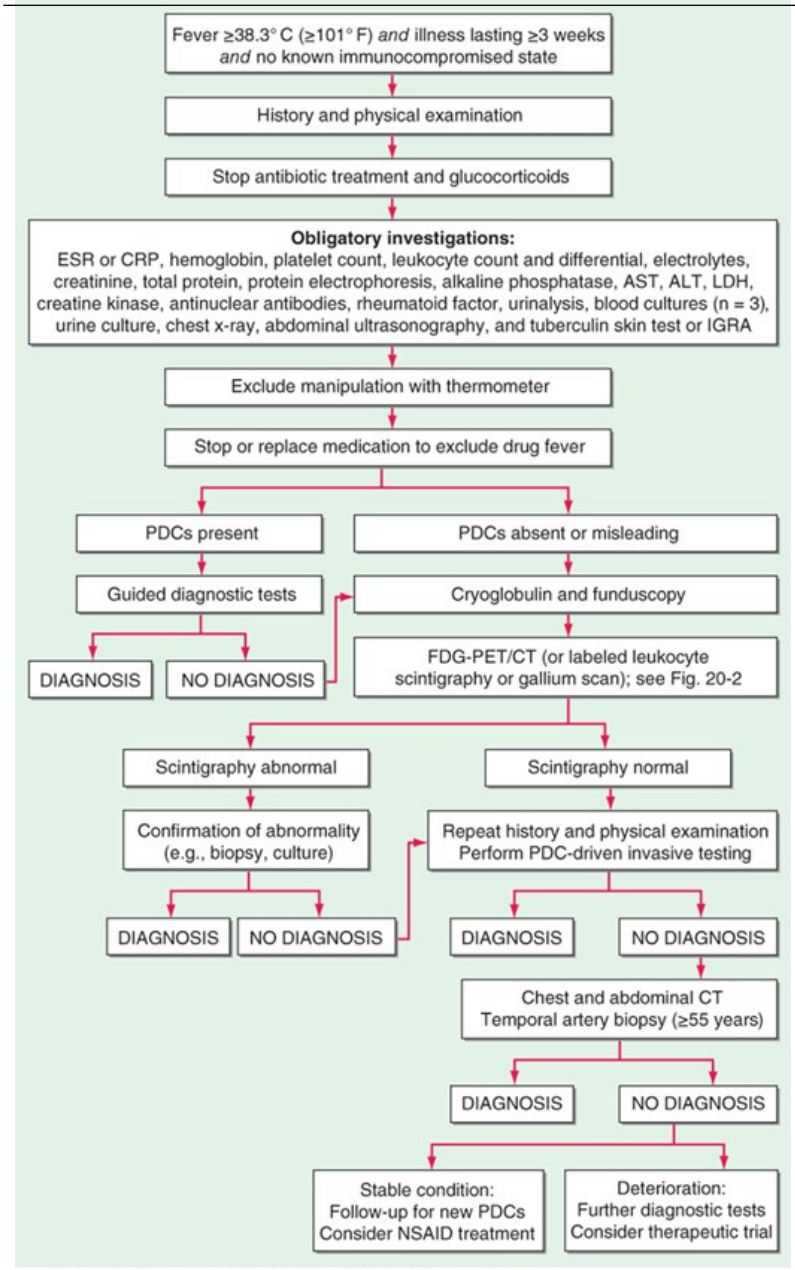
Structured approach to patients with FUO. ALT, alanine aminotransferase; AST, aspartate aminotransferase; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FDG-PET/CT, ^{18}F-fluorodeoxyglucose positron emission tomography combined with low-dose CT; IGRA, interferon γ release assay; LDH, lactate dehydrogenase; NSAID, nonsteroidal anti-inflammatory drug; PDCs, potentially diagnostic clues (all localizing signs, symptoms, and abnormalities potentially pointing toward a diagnosis).



Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

symptoms, and abnormalities potentially pointing toward a diagnosis).



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 20-2

FDG-PET/CT in a patient with FUO. This 72-year-old woman presented with a low-grade fever and severe fatigue of almost 3 months' duration. An extensive history was taken, but the patient had no specific complaints and had not traveled recently. Her previous history was unremarkable, and she did not use any drugs. Physical examination, including palpation of the temporal arteries, yielded completely normal results. Laboratory examination showed normocytic anemia, a C-reactive protein level of 43 mg/L, an erythrocyte sedimentation rate of 87 mm/h, and mild hypoalbuminemia. Results of the other obligatory tests were all normal. Since there were no potentially diagnostic clues, FDG-PET/CT was performed. This test showed increased FDG uptake in all major arteries (carotid, jugular, and subclavian arteries; thoracic and abdominal aorta; iliac, femoral, and popliteal arteries) and in the soft tissue around the shoulders, hips, and knees—findings compatible with large-vessel vasculitis and polymyalgia rheumatica. Within 1 week after the initiation of treatment with prednisone (60 mg once daily), the patient completely recovered. After 1 month, the prednisone dose was slowly tapered. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility




ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

© McGraw Hill Education. All rights reserved.)



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 20-3

All Reported Causes of Recurrent Fever[a]

Infections	
Bacterial, nonspecific	Apical granuloma, diverticulitis, prostatitis, recurrent bacteremia caused by colonic neoplasia or persistent focal infection, recurrent cellulitis, recurrent cholangitis or cholecystitis, recurrent pneumonia, recurrent sinusitis, recurrent urinary tract infection
Bacterial, specific	Bartonellosis, brucellosis, chronic gonococcemia, chronic meningococcemia, louse-borne relapsing fever (*Borrelia recurrentis*), melioidosis (*Pseudomonas pseudomallei*), Q fever (*Coxiella burnetii*), salmonellosis, *Spirillum minor* infection, *Streptobacillus moniliformis* infection, syphilis, tick-borne relapsing fever (*Borrelia duttonii*), tularemia, Whipple's disease (*Tropheryma whipplei*), yersiniosis
Fungal	Coccidioidomycosis, histoplasmosis, paracoccidioidomycosis
Parasitic	Babesiosis, malaria, toxoplasmosis, trypanosomiasis, visceral leishmaniasis
Viral	Cytomegalovirus infection, Epstein-Barr virus infection, herpes simplex
Noninfectious Inflammatory Diseases	
Systemic rheumatic and autoimmune diseases	Ankylosing spondylitis, antiphospholipid syndrome, autoimmune hemolytic anemia, autoimmune hepatitis, Behçet's disease, cryoglobulinemia, gout, polymyositis, pseudogout, reactive arthritis, relapsing polychondritis, systemic lupus erythematosus
Vasculitis	Churg-Strauss syndrome, giant cell vasculitis/polymyalgia rheumatica, hypersensitivity vasculitis, polyarteritis nodosa, urticarial vasculitis
Granulomatous diseases	Idiopathic granulomatous hepatitis, sarcoidosis
Autoinflammatory	Adult-onset Still's disease, Blau syndrome, CANDLE (chronic atypical neutrophilic dermatitis with lipodystrophy and elevated
	PAPA syndrome (pyogenic sterile arthritis, pyoderma gangrenosum, and acne), PFAPA syndrome (periodic fever, aphthous stomatitis, pharyngitis, adenitis), recurrent idiopathic pericarditis, SAPHO (synovitis, acne, pustulosis, hyperostosis, osteomyelitis),

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Autoinflammatory syndromes	Adult-onset Still's disease, Blau syndrome, CANDLE (chronic atypical neutrophilic dermatitis with lipodystrophy and elevated temperature syndrome), CAPS[b] (cryopyrin-associated periodic syndrome), CRMO (chronic recurrent multifocal osteomyelitis), Crohn's disease, DIRA (deficiency of the interleukin 1 receptor antagonist), familial Mediterranean fever, hemophagocytic syndrome, hyper-IgD syndrome (HIDS, also known as mevalonate kinase deficiency), juvenile idiopathic arthritis, NLRC4-activating mutations, PAPA syndrome (pyogenic sterile arthritis, pyoderma gangrenosum, and acne), PFAPA syndrome (periodic fever, aphthous stomatitis, pharyngitis, adenitis), recurrent idiopathic pericarditis, SAPHO (synovitis, acne, pustulosis, hyperostosis, osteomyelitis), SAVI (stimulator of interferon genes [STING]–associated vasculopathy with onset in infancy), Schnitzler syndrome, TRAPS (tumor necrosis factor receptor–associated periodic syndrome)
Neoplasms	
	Angioimmunoblastic lymphoma, Castleman's disease, colon carcinoma, craniopharyngioma, Hodgkin's disease, malignant histiocytosis, mesothelioma, non-Hodgkin's lymphoma
Miscellaneous Causes	
	Adrenal insufficiency, aortic-enteral fistula, aseptic meningitis (Mollaret's syndrome), atrial myxoma, brewer's yeast ingestion, cholesterol emboli, cyclic neutropenia, drug fever, extrinsic allergic alveolitis, Fabry's disease, factitious disease, fraudulent fever, Gaucher disease, hypersensitivity pneumonitis, hypertriglyceridemia, hypothalamic hypopituitarism, inflammatory pseudotumor, metal fume fever, milk protein allergy, polymer fume fever, pulmonary embolism, sclerosing mesenteritis
Thermoregulatory Disorders	
Central	Hypothalamic dysfunction
Peripheral	Anhidrotic ectodermal dysplasia, exercise-induced hyperthermia, pheochromocytoma

[a]This table includes all causes of recurrent fever that have been described in the literature.

[b]CAPS includes chronic infantile neurologic cutaneous and articular syndrome (CINCA, also known as neonatal-onset multisystem inflammatory disease, or NOMID), familial cold autoinflammatory syndrome (FCAS), and Muckle-Wells syndrome.

TREATMENT

Fever of Unknown Origin

Empirical therapeutic trials with antibiotics, glucocorticoids, or antituberculous agents should be avoided in FUO except when a patient's condition is rapidly deteriorating after the aforementioned diagnostic tests have failed to provide a definite diagnosis.

ANTIBIOTICS AND ANTITUBERCULOUS THERAPY

Antibiotic or antituberculous therapy may irrevocably diminish the ability to culture fastidious bacteria or mycobacteria. However, hemodynamic instability or neutropenia is a good indication for empirical antibiotic therapy. If the TST or IGRA is positive or if granulomatous disease is present with anergy and sarcoidosis seems unlikely, a trial of therapy for tuberculosis should be started. Especially in miliary tuberculosis, it may be very difficult to obtain a rapid diagnosis. If the fever does not respond after 6 weeks of empirical antituberculous treatment, another diagnosis should be considered.

COLCHICINE, NONSTEROIDAL ANTI-INFLAMMATORY DRUGS, AND GLUCOCORTICOIDS

Colchicine is highly effective in preventing attacks of familial Mediterranean fever (FMF) but is not always effective once an attack is well under way. When FMF is suspected, the response to colchicine is not a completely reliable diagnostic tool in the acute phase, but with colchicine treatment most patients show remarkable improvements in the frequency and severity of subsequent febrile episodes within weeks to months. Therefore, colchicine may be tried in patients with features compatible with FMF, especially when these patients originate from a high-prevalence region.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The effects of glucocorticoids on giant cell arteritis and polymyalgia rheumatica are equally impressive. Early empirical trials with glucocorticoids,

[a]This table includes all causes of recurrent fever that have been described in the literature.

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Fever of Unknown Origin

Empirical therapeutic trials with antibiotics, glucocorticoids, or antituberculous agents should be avoided in FUO except when a patient's condition is rapidly deteriorating after the aforementioned diagnostic tests have failed to provide a definite diagnosis.

ANTIBIOTICS AND ANTITUBERCULOUS THERAPY

Antibiotic or antituberculous therapy may irrevocably diminish the ability to culture fastidious bacteria or mycobacteria. However, hemodynamic instability or neutropenia is a good indication for empirical antibiotic therapy. If the TST or IGRA is positive or if granulomatous disease is present with anergy and sarcoidosis seems unlikely, a trial of therapy for tuberculosis should be started. Especially in miliary tuberculosis, it may be very difficult to obtain a rapid diagnosis. If the fever does not respond after 6 weeks of empirical antituberculous treatment, another diagnosis should be considered.

COLCHICINE, NONSTEROIDAL ANTI-INFLAMMATORY DRUGS, AND GLUCOCORTICOIDS

Colchicine is highly effective in preventing attacks of familial Mediterranean fever (FMF) but is not always effective once an attack is well under way. When FMF is suspected, the response to colchicine is not a completely reliable diagnostic tool in the acute phase, but with colchicine treatment most patients show remarkable improvements in the frequency and severity of subsequent febrile episodes within weeks to months. Therefore, colchicine may be tried in patients with features compatible with FMF, especially when these patients originate from a high-prevalence region.

If the fever persists and the source remains elusive after completion of the later-stage investigations, supportive treatment with nonsteroidal anti-inflammatory drugs (NSAIDs) can be helpful. The response of adult-onset Still's disease to NSAIDs is dramatic in some cases.

The effects of glucocorticoids on giant cell arteritis and polymyalgia rheumatica are equally impressive. Early empirical trials with glucocorticoids, however, decrease the chances of reaching a diagnosis for which more specific and sometimes life-saving treatment might be more appropriate, such as malignant lymphoma. The ability of NSAIDs and glucocorticoids to mask fever while permitting the spread of infection or lymphoma dictates that their use should be avoided unless infectious diseases and malignant lymphoma have been largely ruled out and inflammatory disease is probable and is likely to be debilitating or threatening.

INTERLEUKIN 1 INHIBITION

Interleukin (IL) 1 is a key cytokine in local and systemic inflammation and the febrile response. The availability of specific IL-1-targeting agents has revealed a pathologic role of IL-1-mediated inflammation in a growing list of diseases. Anakinra, a recombinant form of the naturally occurring IL-1 receptor antagonist (IL-1Ra), blocks the activity of both IL-1α and IL-1β. Anakinra is extremely effective in the treatment of many autoinflammatory syndromes, such as FMF, cryopyrin-associated periodic syndrome, tumor necrosis factor receptor–associated periodic syndrome, mevalonate kinase deficiency (hyper IgD syndrome), Schnitzler syndrome, and adult onset Still's disease. There are many other chronic inflammatory disorders in which anti-IL-1 therapy is highly effective. A therapeutic trial with anakinra can be considered in patients whose FUO has not been diagnosed after later-stage diagnostic tests. Although most chronic inflammatory conditions without a known basis can be controlled with glucocorticoids, monotherapy with IL-1 blockade can provide improved control without the metabolic, immunologic, and gastrointestinal side effects of glucocorticoid administration.

PROGNOSIS

In patients in whom FUO remains unexplained, prognosis is favorable. Two large studies on mortality in these patients have been performed. The first study included 436 patients of whom 168 remained without a diagnosis. Of these, 4 (2.4%) died during follow-up. All 4 patients died during the index admission, and in 2 of them a diagnosis was made upon autopsy (1 had intravascular lymphoma and 1 had bilateral pneumonia). The second study included 131 patients with unexplained FUO. Of these patients, 9 (6.9%) died during a median follow-up of 5 years. In 6 of these patients the cause of death was known, and in 5 of them death was considered unrelated to the febrile disease. Overall, FUO-related mortality rates have continuously declined over recent decades. The majority of fevers are caused by treatable diseases, and the risk of death related to FUO is, of course, dependent on the underlying disease.

FURTHER READING

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

(Baltimore) 86:26, 2007. [PubMed: 17220753]

the underlying disease.

FURTHER READING

Bleeker-Rovers CP et al: A prospective multicenter study on fever of unknown origin: The yield of a structured diagnostic protocol. Medicine (Baltimore) 86:26, 2007. [PubMed: 17220753]

Kouijzer IJE et al: Fever of unknown origin: The value of FDG-PET/CT. Semin Nucl Med 48:100, 2018. [PubMed: 29452615]

Mulders-Manders C et al: Fever of unknown origin. Clin Med 15:280, 2015.

Mulders-Manders C et al: Long-term prognosis, treatment, and outcome of patients with fever of unknown origin in whom no diagnosis was made despite extensive investigation: A questionnaire based study. Medicine (Baltimore) 97:e11241, 2018. [PubMed: 29924054]

Vanderschueren S et al: Inflammation of unknown origin versus fever of unknown origin: Two of a kind. Eur J Intern Med 20:4, 2009.

Vanderschueren S et al: Mortality in patients presenting with fever of unknown origin. Acta Clin Belg 69:12, 2014. [PubMed: 24635393]

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 20: Fever of Unknown Origin, Chantal P. Bleeker-Rovers; Catharina M. Mulders-Manders; Jos W. M. van der Meer
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 22: Dizziness and Vertigo

Mark F. Walker; Robert B. Daroff

INTRODUCTION

Dizziness is an imprecise symptom used to describe a variety of common sensations that include vertigo, light-headedness, faintness, and imbalance. *Vertigo* refers to a sense of spinning or other motion that may be physiological, occurring during or after a sustained head rotation, or pathological, due to vestibular dysfunction. The term *light-headedness* is classically applied to presyncopal sensations resulting from brain hypoperfusion but as used by patients has little specificity, as it may also refer to other symptoms such as disequilibrium and imbalance. A challenge to diagnosis is that patients often have difficulty distinguishing among these various symptoms, and the words they choose do not reliably indicate the underlying etiology.

There are many causes of dizziness. Vestibular dizziness (vertigo or imbalance) may be due to peripheral disorders that affect the labyrinths or vestibular nerves, or it may result from disruption of central vestibular pathways. It may be paroxysmal or due to a fixed unilateral or bilateral vestibular deficit. Acute unilateral lesions cause vertigo due to a sudden imbalance in vestibular inputs from the two labyrinths. Bilateral lesions cause imbalance and instability of vision when the head moves (*oscillopsia*) due to loss of normal vestibular reflexes.

Presyncopal dizziness occurs when cardiac dysrhythmia, orthostatic hypotension, medication effects, or another cause leads to brain hypoperfusion. Such presyncopal sensations vary in duration; they may increase in severity until loss of consciousness occurs, or they may resolve before loss of consciousness if the cerebral ischemia is corrected. Faintness and syncope, which are discussed in detail in **Chap. 21**, should always be considered when one is evaluating patients with brief episodes of dizziness or dizziness that occurs with upright posture. Other causes of dizziness include nonvestibular imbalance, gait disorders (e.g., loss of proprioception from sensory neuropathy, parkinsonism), and anxiety.

When evaluating patients with dizziness, questions to consider include the following: (1) Is it dangerous (e.g., arrhythmia, transient ischemic attack/stroke)? (2) Is it vestibular? (3) If vestibular, is it peripheral or central? A careful history and examination often provide sufficient information to answer these questions and determine whether additional studies or referral to a specialist is necessary.

APPROACH TO THE PATIENT

Dizziness

HISTORY

When a patient presents with dizziness, the first step is to delineate more precisely the nature of the symptom. In the case of vestibular disorders, the physical symptoms depend on whether the lesion is unilateral or bilateral, and whether it is acute or chronic. Vertigo, an illusion of self or environmental motion, implies an acute asymmetry of vestibular inputs from the two labyrinths or in their central pathways. Symmetric bilateral vestibular hypofunction causes imbalance but no vertigo. Because of the ambiguity in patients' descriptions of their symptoms, diagnosis based simply on symptom characteristics is typically unreliable. Thus the history should focus closely on other features, including whether this is the first attack, the duration of this and any prior episodes, provoking factors, and accompanying symptoms.

Dizziness can be divided into episodes that last for seconds, minutes, hours, or days. Common causes of brief dizziness (seconds) include benign paroxysmal positional vertigo (BPPV) and orthostatic hypotension, both of which typically are provoked by changes in head and/or body position relative to gravity. Attacks of vestibular migraine and Ménière's disease often last hours. When episodes are of intermediate duration (minutes), transient ischemic attacks of the posterior circulation should be considered, although migraine and other causes are also possible.

Symptoms that accompany vertigo may be helpful in distinguishing peripheral vestibular lesions from central causes. Unilateral hearing loss and other acute aural symptoms (ear pain, pressure, fullness, new tinnitus) typically point to a peripheral cause. Because the auditory pathways quickly

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 22: Dizziness and Vertigo, Mark F. Walker; Robert B. Daroff

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

EXAMINATION

Because dizziness and imbalance can be a manifestation of a variety of neurologic disorders, the neurologic examination is important in the

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Symptoms that accompany vertigo may be helpful in distinguishing peripheral vestibular lesions from central causes. Unilateral hearing loss and other acute aural symptoms (ear pain, pressure, fullness, new tinnitus) typically point to a peripheral cause. Because the auditory pathways quickly become bilateral upon entering the brainstem, central lesions are unlikely to cause unilateral hearing loss unless the lesion lies near the root entry zone of the auditory nerve. Symptoms such as double vision, numbness, and limb ataxia suggest a brainstem or cerebellar lesion.

EXAMINATION

Because dizziness and imbalance can be a manifestation of a variety of neurologic disorders, the neurologic examination is important in the evaluation of these patients. Focus should be given to assessment of eye movements, vestibular function, and hearing. The range of eye movements and whether they are equal in each eye should be observed. Peripheral eye movement disorders (e.g., cranial neuropathies, eye muscle weakness) are usually disconjugate (different in the two eyes). One should check pursuit (the ability to follow a smoothly moving target) and saccades (the ability to look back and forth accurately between two targets). Poor pursuit or inaccurate (dysmetric) saccades usually indicate central pathology, often involving the cerebellum. Alignment of the two eyes can be checked with a cover test: while the patient is looking at a target, alternately cover the eyes and observe for corrective saccades. A vertical misalignment may indicate a brainstem or cerebellar lesion. Finally, one should look for spontaneous nystagmus, an involuntary back-and-forth movement of the eyes. Nystagmus is most often of the jerk type, in which a slow drift (slow phase) in one direction alternates with a rapid saccadic movement (quick phase or fast phase) in the opposite direction that resets the position of the eyes in the orbits. Except in the case of acute vestibulopathy (e.g., vestibular neuritis), if primary position nystagmus is easily seen in the light, it is probably due to a central cause. Two forms of nystagmus that are characteristic of lesions of the cerebellar pathways are vertical nystagmus with downward fast phases (downbeat nystagmus) and horizontal nystagmus that changes direction with gaze (gaze-evoked nystagmus). By contrast, peripheral lesions typically cause unidirectional horizontal nystagmus. Use of Frenzel eyeglasses (self-illuminated goggles with convex lenses that blur the patient's vision but allow the examiner to see the eyes greatly magnified) or infrared video goggles can aid in the detection of peripheral vestibular nystagmus, because they reduce the patient's ability to use visual fixation to suppress nystagmus. **Table 22-1** outlines key findings that help distinguish peripheral from central causes of vertigo.

The most useful bedside test of peripheral vestibular function is the head impulse test, in which the vestibulo-ocular reflex (VOR) is assessed with small-amplitude (~20 degrees) rapid head rotations. While the patient fixates on a target, the head is rotated quickly to the left or right. If the VOR is deficient, the rotation is followed by a catch-up saccade in the opposite direction (e.g., a leftward saccade after a rightward rotation). The head impulse test can identify both unilateral (catch-up saccades after rotations toward the weak side) and bilateral (catch-up saccades after rotations in both directions) vestibular hypofunction.

All patients with episodic dizziness, especially if provoked by positional change, should be tested with the Dix-Hallpike maneuver. The patient begins in a sitting position with the head turned 45 degrees; holding the back of the head, the examiner then lowers the patient into a supine position with the head extended backward by about 20 degrees while watching the eyes. Posterior canal BPPV can be diagnosed confidently if transient upbeating-torsional nystagmus is seen. If no nystagmus is observed after 15–20 s, the patient is raised to the sitting position, and the procedure is repeated with the head turned to the other side. Again, Frenzel goggles may improve the sensitivity of the test.

Dynamic visual acuity is a functional test that can be useful in assessing vestibular function. Visual acuity is measured with the head still and when the head is rotated back and forth by the examiner (about 1–2 Hz). A drop in visual acuity during head motion of more than one line on a near card or Snellen chart is abnormal and indicates vestibular dysfunction.

ANCILLARY TESTING

The choice of ancillary tests should be guided by the history and examination findings. Audiometry should be performed whenever a vestibular disorder is suspected. Unilateral sensorineural hearing loss supports a peripheral disorder (e.g., vestibular schwannoma). Predominantly low-frequency hearing loss is characteristic of Ménière's disease. Videonystagmography includes recordings of spontaneous nystagmus (if present) and measurement of positional nystagmus. Caloric testing compares the responses of the two horizontal semicircular canals, while video head-impulse testing measures the integrity of each of the six semicircular canals. Vestibular evoked potentials assess otolith reflexes. The test battery often includes recording of saccades and pursuit to evaluate central ocular motor function. Neuroimaging is important if a central vestibular disorder is suspected. In addition, patients with unexplained unilateral hearing loss or vestibular hypofunction should undergo MRI of the internal auditory canals, including administration of gadolinium, to rule out a schwannoma.

TABLE 22-1

Features of Peripheral and Central Vertigo

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 22: Dizziness and Vertigo, Mark F. Walker; Robert B. Daroff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

direction with gaze is due to a central lesion.

- Transient mixed vertical-torsional nystagmus occurs in benign paroxysmal positional vertigo (BPPV), but pure vertical or pure torsional nystagmus is

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 22-1

Features of Peripheral and Central Vertigo

- Nystagmus from an acute peripheral lesion is unidirectional, with fast phases beating away from the ear with the lesion. Nystagmus that changes direction with gaze is due to a central lesion.
- Transient mixed vertical-torsional nystagmus occurs in benign paroxysmal positional vertigo (BPPV), but pure vertical or pure torsional nystagmus is a central sign.
- Nystagmus from a peripheral lesion may be inhibited by visual fixation, whereas central nystagmus is not suppressed.
- Absence of a head impulse sign in a patient with acute prolonged vertigo should suggest a central cause.
- Unilateral hearing loss suggests peripheral vertigo. Findings such as diplopia, dysarthria, and limb ataxia suggest a central disorder.

DIFFERENTIAL DIAGNOSIS AND TREATMENT

Treatment of vestibular symptoms should be driven by the underlying diagnosis. Simply treating dizziness with vestibular suppressant medications is often not helpful and may make the symptoms worse and prolong recovery. The diagnostic and specific treatment approaches for the most commonly encountered vestibular disorders are discussed below.

ACUTE PROLONGED VERTIGO (VESTIBULAR NEURITIS)

An acute unilateral vestibular lesion causes constant vertigo, nausea, vomiting, oscillopsia (motion of the visual scene), and imbalance. These symptoms are due to a sudden asymmetry of inputs from the two labyrinths or in their central connections, simulating a continuous rotation of the head. Unlike BPPV, continuous vertigo persists even when the head remains still.

When a patient presents with an acute vestibular syndrome, the most important question is whether the lesion is central (e.g., a cerebellar or brainstem infarct or hemorrhage), which may be life-threatening, or peripheral, affecting the vestibular nerve or labyrinth (vestibular neuritis). Attention should be given to any symptoms or signs that point to central dysfunction (diplopia, weakness or numbness, dysarthria). The pattern of spontaneous nystagmus, if present, may be helpful **(Table 22-1)**. If the head impulse test is normal, an acute peripheral vestibular lesion is unlikely. A central lesion cannot always be excluded with certainty based on symptoms and examination alone; thus older patients with vascular risk factors who present with an acute vestibular syndrome should be evaluated for the possibility of stroke even when there are no specific findings that indicate a central lesion.

Most patients with vestibular neuritis recover spontaneously, although chronic dizziness, motion sensitivity, and disequilibrium may persist. The role of early glucocorticoid therapy is uncertain, as studies have yielded disparate results. Antiviral medications are of no proven benefit and are not typically given unless there is evidence to suggest herpes zoster oticus (Ramsay Hunt syndrome). Vestibular suppressant medications may reduce acute symptoms but should be avoided after the first several days because they may impede central compensation and recovery. Patients should be encouraged to resume a normal level of activity as soon as possible, and directed vestibular rehabilitation therapy may accelerate improvement.

BENIGN PAROXYSMAL POSITIONAL VERTIGO

BPPV is a common cause of recurrent vertigo. Episodes are brief (<1 min and typically 15–20 s) and are always provoked by changes in head position relative to gravity, such as lying down, rising from a supine position, and extending the head to look upward. Rolling over in bed is a common trigger that may help to distinguish BPPV from orthostatic hypotension. The attacks are caused by free-floating otoconia (calcium carbonate crystals) that have been dislodged from the utricular macula and have moved into one of the semicircular canals, usually the posterior canal. When head position changes, gravity causes the otoconia to move within the canal, producing vertigo and nystagmus. With posterior canal BPPV, the nystagmus beats upward and torsionally (the upper poles of the eyes beat toward the affected lower ear). Less commonly, the otoconia enter the horizontal canal, resulting in a horizontal nystagmus when the patient is lying with either ear down. Superior (also called anterior) canal involvement is rare. BPPV is treated with repositioning maneuvers that use gravity to remove the otoconia from the semicircular canal. For posterior canal BPPV, the Epley maneuver **(Fig. 22-1)** is the most commonly used procedure. For more refractory cases of BPPV, patients can be taught a variant of this maneuver that they can perform alone at home. A demonstration of the Epley maneuver is available online (*http://www.dizziness-and*

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 22: Dizziness and Vertigo, Mark F. Walker; Robert B. Daroff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Modified Epley maneuver for treatment of benign paroxysmal positional vertigo of the right (*top panels*) and left (*bottom panels*) posterior

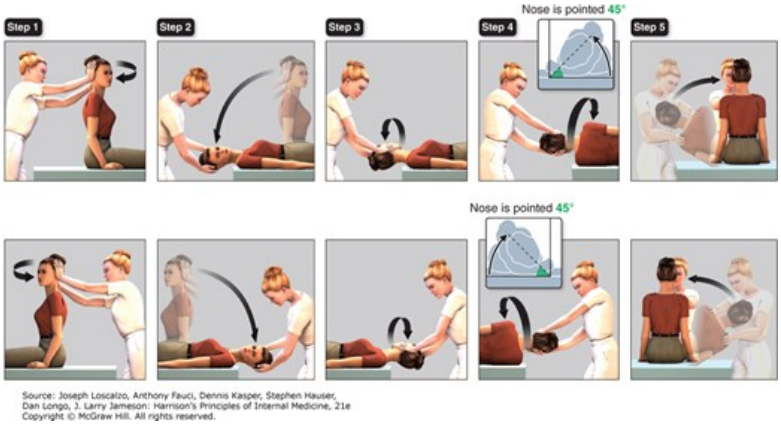
ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

maneuver **(Fig. 22-1)** is the most commonly used procedure. For more refractory cases of BPPV, patients can be taught a variant of this maneuver that they can perform alone at home. A demonstration of the Epley maneuver is available online (*http://www.dizziness-and-balance.com/disorders/bppv/bppv.html*).

FIGURE 22-1

Modified Epley maneuver for treatment of benign paroxysmal positional vertigo of the right (*top panels*) and left (*bottom panels*) posterior semicircular canals. **Step 1.** With the patient seated, turn the head 45 degrees toward the affected ear. **Step 2.** Keeping the head turned, lower the patient to the head-hanging position and hold for at least 30 s and until nystagmus disappears. **Step 3.** Without lifting the head, turn it 90 degrees toward the other side. Hold for another 30 s. **Step 4.** Rotate the patient onto her side while turning the head another 90 degrees, so that the nose is pointed down 45 degrees. Hold again for 30 s. **Step 5.** Have the patient sit up on the side of the table. After a brief rest, the maneuver should be repeated to confirm successful treatment. *(Figure adapted from http://www.dizziness-and-balance.com/disorders/bppv/movies/Epley-480x640.avi.)*



VESTIBULAR MIGRAINE

Vestibular migraine is a common yet underdiagnosed cause of episodic vertigo. Vertigo sometimes precedes a typical migraine headache but more often occurs without headache or with only a mild headache. Some patients who have had frequent migraine headaches in the past present later in life with vestibular migraine as the predominant problem. In vestibular migraine, the duration of vertigo may be from minutes to hours, and some migraineurs also experience more prolonged periods of disequilibrium (lasting days to weeks). Motion sensitivity and sensitivity to visual motion (e.g., movies) are common. Even in the absence of headache, other migraine features may be present, such as photophobia, phonophobia, or a visual aura. Although data from controlled studies are generally lacking, vestibular migraine typically is treated with medications that are used for prophylaxis of migraine headaches **(Chap. 430)**. Antiemetics may be helpful to relieve symptoms at the time of an attack.

MÉNIÈRE'S DISEASE

Attacks of Ménière's disease consist of vertigo and hearing loss, as well as pain, pressure, and/or fullness in the affected ear. Low-frequency hearing loss and aural symptoms are key features that distinguish Ménière's disease from other peripheral vestibulopathies and from vestibular migraine. Audiometry at the time of an attack shows a characteristic asymmetric low-frequency hearing loss; hearing commonly improves between attacks, although permanent hearing loss may eventually occur. Ménière's disease is associated with excess endolymph fluid in the inner ear; hence the term *endolymphatic hydrops.* The exact pathophysiological mechanism, however, remains unclear. Patients suspected of having Ménière's disease should be referred to an otolaryngologist for further evaluation. Diuretics and sodium restriction are typically the initial treatments. If attacks persist, injections of glucocorticoids or gentamicin into the middle ear may be considered. Nonablative surgical options include decompression and shunting of the endolymphatic sac. Full ablative procedures (vestibular nerve section, labyrinthectomy) are seldom required.

VESTIBULAR SCHWANNOMA

Vestibular schwannomas (sometimes termed *acoustic neuromas*) and other tumors at the cerebellopontine angle cause slowly progressive unilateral sensorineural hearing loss and vestibular hypofunction. These patients typically do not have vertigo, because the gradual vestibular deficit is compensated centrally as it develops. The diagnosis often is not made until there is sufficient hearing loss to be noticed. The vestibular examination will show a deficient response to the head impulse test when the head is rotated toward the affected side, but nystagmus will not be prominent. As

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 22: Dizziness and Vertigo, Mark F. Walker; Robert B. Daroff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

BILATERAL VESTIBULAR HYPOFUNCTION

compensated centrally as it develops. The diagnosis often is not made until there is sufficient hearing loss to be noticed. The vestibular examination will show a deficient response to the head impulse test when the head is rotated toward the affected side, but nystagmus will not be prominent. As noted above, patients with unexplained unilateral sensorineural hearing loss or vestibular hypofunction require MRI of the internal auditory canals to look for a schwannoma.

BILATERAL VESTIBULAR HYPOFUNCTION

Patients with bilateral loss of vestibular function also typically do not have vertigo, because vestibular function is lost on both sides simultaneously, and there is no asymmetry of vestibular input. Symptoms include loss of balance, particularly in the dark, where vestibular input is most critical, and oscillopsia during head movement, such as while walking or riding in a car. Bilateral vestibular hypofunction may be (1) idiopathic and progressive, (2) part of a neurodegenerative disorder, or (3) iatrogenic due to medication ototoxicity (most commonly gentamicin or other aminoglycoside antibiotics). Other causes include bilateral vestibular schwannomas (neurofibromatosis type 2), autoimmune disease, superficial siderosis, and meningeal-based infection or tumor. It also may occur in patients with peripheral polyneuropathy; in these patients, both vestibular loss and impaired proprioception may contribute to poor balance. Finally, unilateral processes such as vestibular neuritis and Ménière's disease may involve both ears sequentially, resulting in bilateral vestibulopathy.

Examination findings include diminished *dynamic visual acuity* (see above) due to loss of stable vision when the head is moving, abnormal head impulse responses in both directions, and a Romberg sign. Responses to caloric testing are reduced. Patients with bilateral vestibular hypofunction should be referred for vestibular rehabilitation therapy. Vestibular suppressant medications should not be used, as they will increase the imbalance. Evaluation by a neurologist is important not only to confirm the diagnosis but also to consider any other associated neurologic abnormalities that may clarify the etiology.

CENTRAL VESTIBULAR DISORDERS

Central lesions causing vertigo typically involve vestibular pathways in the brainstem and/or cerebellum. They may be due to discrete lesions, such as from ischemic or hemorrhagic stroke **(Chaps. 426–428)**, demyelination **(Chap. 444)**, or tumors **(Chap. 90)**, or they may be due to neurodegenerative conditions that include the vestibulocerebellum **(Chaps. 431–434)**. Subacute cerebellar degeneration may be due to immune, including paraneoplastic, processes **(Chaps. 94 and 439)**. **Table 22-1** outlines important features of the history and examination that help to identify central vestibular disorders. Acute central vertigo is a medical emergency, due to the possibility of life-threatening stroke or hemorrhage. All patients with suspected central vestibular disorders should undergo brain MRI, and the patient should be referred for full neurologic evaluation.

PSYCHOSOMATIC AND FUNCTIONAL DIZZINESS

Psychological factors play an important role in chronic dizziness. First, dizziness may be a somatic manifestation of a psychiatric condition such as major depression, anxiety, or panic disorder **(Chap. 452)**. Second, patients may develop anxiety and autonomic symptoms as a consequence or comorbidity of an independent vestibular disorder. One particular form of this has been termed variously *phobic postural vertigo*, *psychophysiologic vertigo*, or *chronic subjective dizziness*, but is now referred to as *persistent postural-perceptual dizziness (PPPD)*. These patients have a chronic feeling (3 months or longer) of fluctuating dizziness and disequilibrium that is present at rest but worse while standing. There is an increased sensitivity to self-motion and visual motion (e.g., watching movies), and a particular intensification of symptoms when moving through complex visual environments such as supermarkets. Although there may be a past history of an acute vestibular disorder (e.g., vestibular neuritis), the neuro-otologic examination and vestibular testing are normal or indicative of a compensated vestibular deficit, indicating that the ongoing subjective dizziness cannot be explained by a primary vestibular pathology. Anxiety disorders are particularly common in patients with chronic dizziness; when present, they contribute substantially to the morbidity. Treatment approaches for PPPD include pharmacological therapy with selective serotonin reuptake inhibitors (SSRIs), cognitive-behavioral psychotherapy, and vestibular rehabilitation. Vestibular suppressant medications generally should be avoided.

TREATMENT

Vertigo

Table 22-2 provides a list of commonly used medications for suppression of vertigo. As noted, these medications should be reserved for short-term control of active vertigo, such as during the first few days of acute vestibular neuritis, or for acute attacks of Ménière's disease. They are less helpful for chronic dizziness and, as previously stated, may hinder central compensation. An exception is that benzodiazepines may attenuate psychosomatic dizziness and the associated anxiety, although SSRIs are generally preferable in such patients.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

gaze stabilization and balance.

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

TREATMENT

Vertigo

Table 22-2 provides a list of commonly used medications for suppression of vertigo. As noted, these medications should be reserved for short-term control of active vertigo, such as during the first few days of acute vestibular neuritis, or for acute attacks of Ménière's disease. They are less helpful for chronic dizziness and, as previously stated, may hinder central compensation. An exception is that benzodiazepines may attenuate psychosomatic dizziness and the associated anxiety, although SSRIs are generally preferable in such patients.

Vestibular rehabilitation therapy promotes central adaptation processes that compensate for vestibular loss and also may help habituate motion sensitivity and other symptoms of psychosomatic dizziness. The general approach is to use a graded series of exercises that progressively challenge gaze stabilization and balance.

TABLE 22-2

Treatment of Vertigo

AGENT[a]	DOSE[b]
Antihistamines	
Meclizine	25–50 mg 3 times daily
Dimenhydrinate	50 mg 1–2 times daily
Promethazine	25 mg 2–3 times daily (also can be given rectally and IM)
Benzodiazepines	
Diazepam	2.5 mg 1–3 times daily
Clonazepam	0.25 mg 1–3 times daily
Anticholinergic	
Scopolamine transdermal[c]	Patch
Physical therapy	
Repositioning maneuvers[d]	
Vestibular rehabilitation	
Other	
Diuretics and/or low-sodium (1000 mg/d) diet[e]	
Antimigrainous drugs[f]	
Selective serotonin reuptake inhibitors[g]	

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 22: Dizziness and Vertigo, Mark F. Walker; Robert B. Daroff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

[b] Usual oral (unless otherwise stated) starting dose in adults; a higher maintenance dose can be reached by a gradual increase.

AGENT[a]	DOSE[b]
Antihistamines	
Meclizine	25–50 mg 3 times daily
Dimenhydrinate	50 mg 1–2 times daily
Promethazine	25 mg 2–3 times daily (also can be given rectally and IM)
Benzodiazepines	
Diazepam	2.5 mg 1–3 times daily
Clonazepam	0.25 mg 1–3 times daily
Anticholinergic	
Scopolamine transdermal[c]	Patch
Physical therapy	
Repositioning maneuvers[d]	
Vestibular rehabilitation	
Other	
Diuretics and/or low-sodium (1000 mg/d) diet[e]	
Antimigrainous drugs[f]	
Selective serotonin reuptake inhibitors[g]	

[a]All listed drugs are approved by the US Food and Drug Administration, but most are not approved for the treatment of vertigo.

[b]Usual oral (unless otherwise stated) starting dose in adults; a higher maintenance dose can be reached by a gradual increase.

[c]For motion sickness only.

[d]For benign paroxysmal positional vertigo.

[e]For Ménière's disease.

[f]For vestibular migraine.

[g]For persistent postural-perceptual vertigo and anxiety.

FURTHER READING

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 22: Dizziness and Vertigo, Mark F. Walker; Robert B. Daroff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Huang TC et al: Vestibular migraine: An update on current understanding and future directions. Cephalalgia 40:107, 2020. [PubMed: 31394919]

FURTHER READING

Altissimi G et al: Drugs inducing hearing loss, tinnitus, dizziness and vertigo: An updated guide. Eur Rev Med Pharmacol Sci 24:7946, 2020. [PubMed: 32767320]

Huang TC et al: Vestibular migraine: An update on current understanding and future directions. Cephalalgia 40:107, 2020. [PubMed: 31394919]

Kim JS, Zee DS: Benign paroxysmal positional vertigo. N Engl J Med 370:1138, 2014. [PubMed: 24645946]

Popkirov S et al: Persistent postural-perceptual dizziness (PPPD): a common, characteristic and treatable cause of chronic dizziness. Pract Neurol 18:5, 2018. [PubMed: 29208729]

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 22: Dizziness and Vertigo, Mark F. Walker; Robert B. Daroff
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 23: Fatigue

Jeffrey M. Gelfand; Vanja C. Douglas

INTRODUCTION

Fatigue is one of the most common symptoms in clinical medicine. It is a prominent manifestation of a number of systemic, neurologic, and psychiatric syndromes, although a precise cause will not be identified in a substantial minority of patients. Fatigue refers to the subjective experience of physical and mental weariness, sluggishness, low energy, and exhaustion. In the context of clinical medicine, fatigue is most practically defined as difficulty initiating or maintaining voluntary mental or physical activity. Nearly everyone who has ever been ill with a self-limited infection has experienced this near-universal symptom, and fatigue is usually brought to medical attention only when it is either of unclear cause, fails to remit, or the severity is out of proportion with what would be expected for the associated trigger.

Fatigue should be distinguished from *muscle weakness*, a reduction of neuromuscular power **(Chap. 24)**; most patients complaining of fatigue are not truly weak when direct muscle power is tested. Fatigue is also distinct from *somnolence*, which refers to sleepiness in the context of disturbed sleep-wake physiology **(Chap. 31)**, and from *dyspnea on exertion*, although patients may use the word fatigue to describe any of these symptoms. The task facing clinicians when a patient presents with fatigue is to identify the underlying cause and develop a therapeutic alliance, the goal of which is to spare patients expensive and fruitless diagnostic workups and steer them toward effective therapy.

EPIDEMIOLOGY AND GLOBAL CONSIDERATIONS

Variability in the definitions of fatigue and the survey instruments used in different studies makes it difficult to arrive at precise figures about the global burden of fatigue. The point prevalence of fatigue was 6.7% and the lifetime prevalence was 25% in a large National Institute of Mental Health survey of the U.S. general population. In primary care clinics in Europe and the United States, between 10 and 25% of patients surveyed endorsed symptoms of prolonged (present for >1 month) or chronic (present for >6 months) fatigue, but in only a minority was fatigue the primary reason for seeking medical attention. In a community survey of women in India, 12% reported chronic fatigue. By contrast, the prevalence of chronic fatigue syndrome **(Chap. 450)**, as defined by the U.S. Centers for Disease Control and Prevention, is low.

DIFFERENTIAL DIAGNOSIS

Psychiatric Disease

Fatigue is a common somatic manifestation of many major psychiatric syndromes, including depression, anxiety, and somatoform disorders **(Chap. 452)**. Psychiatric symptoms are reported in more than three-quarters of patients with unexplained chronic fatigue. Even in patients with systemic or neurologic disorders in which fatigue is independently recognized as a symptom, comorbid psychiatric disease may still be an important contributor.

Neurologic Disease

Patients complaining of fatigue often say they feel weak, but upon careful examination, objective muscle weakness is rarely discernible. If found, muscle weakness must then be localized to the central nervous system, peripheral nervous system, neuromuscular junction, or muscle, and appropriate follow-up studies obtained **(Chap. 24)**. *Fatigability* of muscle power is a cardinal manifestation of some neuromuscular disorders such as myasthenia gravis and is distinguished from *fatigue* by finding clinically evident diminution of the amount of force that a muscle generates upon repeated contraction **(Chap. 448)**. Fatigue is one of the most common and bothersome symptoms reported in multiple sclerosis (MS) **(Chap. 444)**, affecting nearly 90% of patients; fatigue in MS can persist between MS attacks and does not necessarily correlate with magnetic resonance imaging (MRI) disease activity. Fatigue is also increasingly identified as a troublesome feature of many neurodegenerative diseases, including Parkinson's disease **(Chap. 435)**, amyotrophic lateral sclerosis **(Chap. 437)**, and central nervous system dysautonomias **(Chap. 440)**. Fatigue after stroke

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 23: Fatigue, Jeffrey M. Gelfand; Vanja C. Douglas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Sleep Disorders

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

disease (Chap. 435), amyotrophic lateral sclerosis (Chap. 437), and central nervous system dysautonomias (Chap. 440). Fatigue after stroke (Chap. 426) is a well-described but poorly understood entity with a widely varying prevalence. Episodic fatigue can be a premonitory symptom of migraine (Chap. 430). Fatigue is also a frequent consequence of traumatic brain injury (Chap. 443), often occurring in association with depression and sleep disorders.

Sleep Disorders

Obstructive sleep apnea is an important cause of excessive daytime sleepiness in association with fatigue and should be investigated using overnight polysomnography, particularly in those with prominent snoring, obesity, or other predictors of obstructive sleep apnea (Chap. 297). Whether the cumulative sleep deprivation that is common in modern society contributes to clinically apparent fatigue is not known (Chap. 31).

Endocrine Disorders

Fatigue, sometimes in association with true muscle weakness, can be a heralding symptom of hypothyroidism (Chap. 383), particularly in the context of hair loss, dry skin, cold intolerance, constipation, and weight gain. Fatigue associated with heat intolerance, sweating, and palpitations is typical of hyperthyroidism (Chap. 384). Adrenal insufficiency (Chap. 386) can also manifest with unexplained fatigue as a primary or prominent symptom, often with anorexia, weight loss, nausea, myalgias, and arthralgias; hyponatremia, hyperkalemia, and hyperpigmentation may be present at time of diagnosis. Mild hypercalcemia can cause fatigue, which may be relatively vague, whereas severe hypercalcemia can lead to lethargy, stupor, and coma (Chap. 410). Both hypoglycemia and hyperglycemia can cause lethargy, often in association with confusion; diabetes mellitus, and in particular type 1 diabetes, is also associated with fatigue independent of glucose levels (Chap. 403). Fatigue may also accompany Cushing's disease, hypoaldosteronism, and hypogonadism. Low vitamin D status has also been associated with fatigue.

Liver and Kidney Disease

Both chronic liver failure and chronic kidney disease can cause fatigue. Over 80% of hemodialysis patients complain of fatigue, which makes it one of the most common symptoms reported by patients in chronic kidney disease (Chap. 311).

Obesity

Obesity (Chap. 401) is associated with fatigue and sleepiness independent of the presence of obstructive sleep apnea. Obese patients undergoing bariatric surgery experience improvement in daytime sleepiness sooner than would be expected if the improvement were solely the result of weight loss and resolution of sleep apnea. A number of other factors common in obese patients are likely contributors as well, including physical inactivity, diabetes, and depression.

Physical Inactivity

Physical inactivity is associated with fatigue, and increasing physical activity can improve fatigue in some patients.

Malnutrition

Although fatigue can be a presenting feature of malnutrition (Chap. 334), nutritional status may also be an important comorbidity and contributor to fatigue in other chronic illnesses, including cancer-associated fatigue.

Infection

Both acute and chronic infections commonly lead to fatigue as part of the broader infectious syndrome. Evaluation for undiagnosed infection as the cause of unexplained fatigue, and particularly prolonged or chronic fatigue, should be guided by the history, physical examination, and infectious risk factors, with particular attention to risk for tuberculosis, HIV, chronic hepatitis, and endocarditis. Infectious mononucleosis may cause prolonged fatigue that persists for weeks to months following the acute illness, but infection with the Epstein-Barr virus is only very rarely the cause of unexplained chronic fatigue. Postinfectious fatigue may also occur following a variety of acute infections. For example, a substantial minority of patients who have recovered from severe acute respiratory syndrome coronavirus (SARS-CoV)-1, SARS-CoV-2, and Ebola virus complain of persistent fatigue.

Drugs

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 23: Fatigue, Jeffrey M. Gelfand; Vanja C. Douglas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Cardiovascular and Pulmonary Disorders

Drugs

Many medications, drugs, drug withdrawal, and chronic alcohol use can all lead to fatigue. Medications that are more likely to be causative include antidepressants, antipsychotics, anxiolytics, opiates, antispasticity agents, antiseizure agents, and beta blockers.

Cardiovascular and Pulmonary Disorders

Fatigue is one of the most taxing symptoms reported by patients with congestive heart failure and chronic obstructive pulmonary disease and negatively affects quality of life. In a population-based cohort study in Norfolk, United Kingdom, fatigue was associated with an increased hazard of all-cause mortality in the general population, but particularly for deaths related to cardiovascular disease.

Malignancy

Fatigue, particularly in association with unexplained weight loss, can be a sign of occult malignancy, but cancer is rarely identified in patients with unexplained chronic fatigue in the absence of other telltale signs or symptoms. Cancer-related fatigue is experienced by 40% of patients at the time of diagnosis and by >80% at some time in the disease course.

Hematologic Disorders

Chronic or progressive anemia may present with fatigue, sometimes in association with exertional tachycardia and breathlessness. Anemia may also contribute to fatigue in chronic illness. Low serum ferritin in the absence of anemia may also cause fatigue that is reversible with iron replacement.

Immune-Mediated Disorders

Fatigue is a prominent complaint in many chronic inflammatory disorders, including systemic lupus erythematosus, polymyalgia rheumatica, rheumatoid arthritis, inflammatory bowel disease, antineutrophil cytoplasmic antibody (ANCA)–associated vasculitis, sarcoidosis, and Sjögren's syndrome, but is not usually an isolated symptom. Fatigue is also associated with primary immunodeficiency diseases.

Pregnancy

Fatigue is very commonly reported by women during all stages of pregnancy and postpartum.

Disorders of Unclear Cause

Myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) **(Chap. 450)** and fibromyalgia **(Chap. 373)** incorporate chronic fatigue as part of the syndromic definition when fatigue is present in association with other criteria, as discussed in the respective chapters. Chronic multisymptom illness, also known as Gulf-War syndrome, is another symptom complex with prominent fatigue; it is most commonly, although not exclusively, observed in veterans of the 1991 Gulf War conflict **(Chap. S7)**. Idiopathic chronic fatigue is used to describe the syndrome of unexplained chronic fatigue in the absence of enough additional clinical features to meet the diagnostic criteria for ME/CFS.

APPROACH TO THE PATIENT

Fatigue

A detailed history focusing on the quality, pattern, time course, associated symptoms, and alleviating factors of fatigue is necessary to define the syndrome and help direct further evaluation and treatment. It is important to determine if fatigue is the appropriate designation, whether symptoms are acute or chronic, and if the impairment is primarily mental, physical, or a combination of the two. The review of systems should attempt to distinguish fatigue from excessive sleepiness, dyspnea on exertion, exercise intolerance, and muscle weakness. The presence of fever, chills, night sweats, or weight loss should raise suspicion for an occult infection or malignancy. A careful review of prescription, over-the-counter, herbal, and recreational drug and alcohol use is required. Circumstances surrounding the onset of symptoms and potential triggers should be investigated. The social history is important, with attention paid to life stressors and adverse experiences, workhours, the social support network, and domestic affairs including a screen for intimate partner violence. Sleep habits and sleep hygiene should be questioned. The impact of fatigue on daily functioning is important to understand the patient's experience and gauge recovery and the success of treatment.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 23: Fatigue, Jeffrey M. Gelfand; Vanja C. Douglas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

objective muscle weakness is present. This is usually a straightforward exercise, although occasionally patients with fatigue have difficulty sustaining effort against resistance and sometimes report that generating full power requires substantial mental effort. On confrontational testing,

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

APPROACH TO THE PATIENT

Fatigue

A detailed history focusing on the quality, pattern, time course, associated symptoms, and alleviating factors of fatigue is necessary to define the syndrome and help direct further evaluation and treatment. It is important to determine if fatigue is the appropriate designation, whether symptoms are acute or chronic, and if the impairment is primarily mental, physical, or a combination of the two. The review of systems should attempt to distinguish fatigue from excessive sleepiness, dyspnea on exertion, exercise intolerance, and muscle weakness. The presence of fever, chills, night sweats, or weight loss should raise suspicion for an occult infection or malignancy. A careful review of prescription, over-the-counter, herbal, and recreational drug and alcohol use is required. Circumstances surrounding the onset of symptoms and potential triggers should be investigated. The social history is important, with attention paid to life stressors and adverse experiences, workhours, the social support network, and domestic affairs including a screen for intimate partner violence. Sleep habits and sleep hygiene should be questioned. The impact of fatigue on daily functioning is important to understand the patient's experience and gauge recovery and the success of treatment.

The physical examination of patients with fatigue is guided by the history and differential diagnosis. A detailed mental status examination should be performed with particular attention to symptoms of depression and anxiety. A formal neurologic examination is required to determine whether objective muscle weakness is present. This is usually a straightforward exercise, although occasionally patients with fatigue have difficulty sustaining effort against resistance and sometimes report that generating full power requires substantial mental effort. On confrontational testing, full power may be generated for only a brief period before the patient suddenly gives way to the examiner. This type of weakness is often referred to as *breakaway weakness* and may or may not be associated with pain. This is contrasted with weakness due to lesions in the motor tracts or lower motor unit, in which the patient's resistance can be overcome in a smooth and steady fashion and full power can never be generated. Occasionally, a patient may demonstrate fatigable weakness, in which power is full when first tested but becomes weak upon repeat evaluation without interval rest. Fatigable weakness, which usually indicates a problem of neuromuscular transmission, never has the sudden breakaway quality that one occasionally observes in patients with fatigue. If the presence or absence of muscle weakness cannot be determined with the physical examination, electromyography with nerve conductions studies can be a helpful ancillary test.

The general physical examination should screen for signs of cardiopulmonary disease, malignancy, lymphadenopathy, organomegaly, infection, liver failure, kidney disease, malnutrition, endocrine abnormalities, and connective tissue disease. In patients with associated widespread musculoskeletal pain, assessment of tender points may help to reveal fibromyalgia. Although the diagnostic yield of the general physical examination may be relatively low in the context of evaluation of unexplained chronic fatigue, elucidating the cause of only 2% of cases in one prospective analysis, the yield of a detailed neuropsychiatric and mental status evaluation is likely to be much higher, revealing a potential explanation for fatigue in up to 75–80% of patients in some series. Furthermore, a complete physical examination demonstrates a serious and systematic approach to the patient's complaint and helps build trust and a therapeutic alliance.

Laboratory testing is likely to identify the cause of chronic fatigue in only about 5% of cases. Beyond a few standard screening tests, laboratory evaluation should be guided by the history and physical examination; extensive testing is likely to lead to incidental findings that require explanation and unnecessary follow-up investigation, and should be avoided in lieu of frequent clinical follow-up. A reasonable approach to screening includes a complete blood count with differential (to screen for anemia, infection, and malignancy), electrolytes (including sodium, potassium, and calcium), glucose, renal function, liver function, and thyroid function. Testing for HIV and adrenal function can also be considered. Published guidelines for chronic fatigue syndrome also recommend an erythrocyte sedimentation rate (ESR) as part of the evaluation for mimics, but unless the value is very high, such nonspecific testing in the absence of other features is unlikely to clarify the situation. Routine screening with an antinuclear antibody (ANA) test is also unlikely to be informative in isolation and is frequently positive at low titers in otherwise healthy adults. Additional unfocused studies, such as whole-body imaging scans, are usually not indicated; in addition to their inconvenience, potential risk, and cost, they often reveal unrelated incidental findings that can prolong the workup unnecessarily.

TREATMENT

Fatigue

The first priority is to address the underlying disorder or disorders that account for fatigue, because this can be curative in select contexts and multimodal approach. However, antidepressants can also cause fatigue and should be discontinued if they are not clearly effective. Cognitive-behavioral therapy has also been demonstrated to be helpful in ME/CFS as well as cancer-associated fatigue. Both cognitive-behavioral therapy and

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 23: Fatigue, Jeffrey M. Gelfand; Vanja C. Douglas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Fatigue

The first priority is to address the underlying disorder or disorders that account for fatigue, because this can be curative in select contexts and palliative in others. Unfortunately, in many chronic illnesses, fatigue may be refractory to traditional disease-modifying therapies, but it is nevertheless important in such cases to evaluate for other potential contributors because the cause may be multifactorial. Antidepressants **(Chap. 452)** may be helpful for treatment of chronic fatigue when symptoms of depression are present and are generally most effective as part of a multimodal approach. However, antidepressants can also cause fatigue and should be discontinued if they are not clearly effective. Cognitive-behavioral therapy has also been demonstrated to be helpful in ME/CFS as well as cancer-associated fatigue. Both cognitive-behavioral therapy and graded exercise therapy, in which physical exercise, most typically walking, is gradually increased with attention to target heart rates to avoid overexertion, were shown to modestly improve walking times and self-reported fatigue measures when compared to standard medical care in patients in the United Kingdom with chronic fatigue. These benefits were maintained after a median follow-up of 2.5 years. Psychostimulants such as amphetamines, modafinil, and armodafinil can help increase alertness and concentration and reduce excessive daytime sleepiness in certain clinical contexts, which may in turn help with symptoms of fatigue in a minority of patients, but they have generally proven to be unhelpful in randomized trials for treating fatigue in posttraumatic brain injury, Parkinson's disease, cancer, and MS. In patients with low vitamin D status, vitamin D replacement may lead to improvement in fatigue.

Development of more effective therapy for fatigue is hampered by limited knowledge of the biologic basis of this symptom, including how fatigue is detected and registered in the nervous system. Proinflammatory cytokines, such as interleukin 1α and 1β and tumor necrosis factor α, might mediate fatigue in some patients. While preliminary studies of biologic therapies that inhibit cytokines have suggested a benefit against fatigue in some patients with inflammatory conditions, this approach has largely not led to improvement in clinical trials that focused on fatigue as the primary endpoint. Nonetheless, specific targeting with cytokine antagonists could represent a possible future approach for some patients.

PROGNOSIS

Acute fatigue significant enough to require medical evaluation is more likely to lead to an identifiable medical, neurologic, or psychiatric cause than is unexplained chronic fatigue. Evaluation of unexplained chronic fatigue most commonly leads to diagnosis of a psychiatric condition or remains unexplained. Identification of a previously undiagnosed serious or life-threatening culprit etiology is rare, even with longitudinal follow-up of patients with unexplained chronic fatigue. Complete resolution is uncommon, at least over the short term, but multidisciplinary treatment approaches can lead to symptomatic improvements that substantially improve quality of life.

FURTHER READING

Basu N et al: Fatigue is associated with excess mortality in the general population: Results from the EPIC-Norfolk study. BMC Med 14:122, 2016. [PubMed: 27543008]

Dukes JC et al: Approach to fatigue: Best practice. Med Clin North Am 105:137, 2021. [PubMed: 33246515]

Roerink ME et al: Interleukin-1 as a mediator of fatigue in disease: A narrative review. J Neuroinflammation 14:16, 2017. [PubMed: 28109186]

Sharpe M et al: Rehabilitative treatments for chronic fatigue syndrome: Long-term follow-up from the PACE trial. Lancet Psychiatry 2:1067, 2015. [PubMed: 26521770]

White PD et al: Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): A randomised trial. Lancet 377:823, 2011. [PubMed: 21334061]

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 23: Fatigue, Jeffrey M. Gelfand; Vanja C. Douglas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 24: Neurologic Causes of Weakness and Paralysis

Stephen L. Hauser

INTRODUCTION

Normal motor function involves integrated muscle activity that is modulated by the activity of the cerebral cortex, basal ganglia, cerebellum, red nucleus, brainstem reticular formation, lateral vestibular nucleus, and spinal cord. Motor system dysfunction leads to weakness or paralysis, discussed in this chapter, or to ataxia **(Chap. 439)** or abnormal movements **(Chap. 436)**. *Weakness* is a reduction in the power that can be exerted by one or more muscles. It must be distinguished from increased *fatigability* (i.e., the inability to sustain the performance of an activity that should be normal for a person of the same age, sex, and size), limitation in function due to pain or articular stiffness, or impaired motor activity because severe *proprioceptive sensory loss* prevents adequate feedback information about the direction and power of movements. It is also distinct from *bradykinesia* (in which increased time is required for full power to be exerted) and *apraxia,* a disorder of planning and initiating a skilled or learned movement unrelated to a significant motor or sensory deficit **(Chap. 30)**.

Paralysis or the suffix "-plegia" indicates weakness so severe that a muscle cannot be contracted at all, whereas *paresis* refers to less severe weakness. The prefix "hemi-" refers to one-half of the body, "para-" to both legs, and "quadri-" to all four limbs.

The *distribution* of weakness helps to localize the underlying lesion. Weakness from involvement of upper motor neurons occurs particularly in the extensors and abductors of the upper limb and the flexors of the lower limb. Lower motor neuron weakness depends on whether involvement is at the level of the anterior horn cells, nerve root, limb plexus, or peripheral nerve—only muscles supplied by the affected structure are weak. Myopathic weakness is generally most marked in proximal muscles. Weakness from impaired neuromuscular transmission has no specific pattern of involvement.

Weakness often is accompanied by other neurologic abnormalities that help indicate the site of the responsible lesion **(Table 24-1)**.

TABLE 24-1

Signs That Distinguish the Origin of Weakness

SIGN	UPPER MOTOR NEURON	LOWER MOTOR NEURON	MYOPATHIC	PSYCHOGENIC
Atrophy	None	Severe	Mild	None
Fasciculations	None	Common	None	None
Tone	Spastic	Decreased	Normal/decreased	Variable/paratonia
Distribution of weakness	Pyramidal/regional	Distal/segmental	Proximal	Variable/inconsistent with daily activities
Muscle stretch reflexes	Hyperactive	Hypoactive/absent	Normal/hypoactive	Normal
Babinski sign	Present	Absent	Absent	Absent

Tone is the resistance of a muscle to passive stretch. Increased tone may be of several types. *Spasticity* is the increase in tone associated with disease of upper motor neurons. It is velocity dependent, has a sudden release after reaching a maximum (the "clasp-knife" phenomenon), and predominantly

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 24: Neurologic Causes of Weakness and Paralysis, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

gegennalten) is increased tone that varies irregularly in a manner seemingly related to the degree of relaxation, is present throughout the range of motion, and affects flexors and extensors equally; it usually results from disease of the frontal lobes. Weakness with *decreased tone (flaccidity)* or

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

of upper motor neurons. It is velocity dependent, has a sudden release after reaching a maximum (the "clasp-knife" phenomenon), and predominantly affects the antigravity muscles (i.e., upper-limb flexors and lower-limb extensors). *Rigidity* is hypertonia that is present throughout the range of motion (a "lead pipe" or "plastic" stiffness) and affects flexors and extensors equally; it sometimes has a cogwheel quality that is enhanced by voluntary movement of the contralateral limb (reinforcement). Rigidity occurs with certain extrapyramidal disorders, such as Parkinson's disease. *Paratonia* (or *gegenhalten*) is increased tone that varies irregularly in a manner seemingly related to the degree of relaxation, is present throughout the range of motion, and affects flexors and extensors equally; it usually results from disease of the frontal lobes. Weakness with *decreased tone (flaccidity)* or normal tone occurs with disorders of *motor units.* A motor unit consists of a single lower motor neuron and all the muscle fibers that it innervates.

Muscle bulk generally is not affected by upper motor neuron lesions, although mild disuse atrophy eventually may occur. By contrast, atrophy is often conspicuous when a lower motor neuron lesion is responsible for weakness and also may occur with advanced muscle disease.

Muscle stretch (tendon) reflexes are usually increased with upper motor neuron lesions but may be decreased or absent for a variable period immediately after onset of an acute lesion. Hyperreflexia is usually—but not invariably—accompanied by loss of *cutaneous reflexes* (such as superficial abdominals; **Chap. 422**) and, in particular, by an extensor plantar (Babinski) response. The muscle stretch reflexes are depressed with lower motor neuron lesions directly involving specific reflex arcs. They generally are preserved in patients with myopathic weakness except in advanced stages, when they sometimes are attenuated. In disorders of the neuromuscular junction, reflex responses may be affected by preceding voluntary activity of affected muscles; such activity may lead to enhancement of initially depressed reflexes in Lambert-Eaton myasthenic syndrome and, conversely, to depression of initially normal reflexes in myasthenia gravis **(Chap. 448)**.

The distinction of *neuropathic* (lower motor neuron) from *myopathic* weakness is sometimes difficult clinically, although distal weakness is likely to be neuropathic, and symmetric proximal weakness myopathic. *Fasciculations* (visible or palpable twitches within a muscle due to the spontaneous discharge of a motor unit) and early atrophy indicate that weakness is neuropathic.

PATHOGENESIS

Upper Motor Neuron Weakness

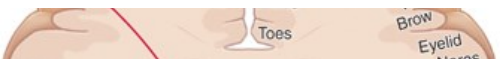
Lesions of the upper motor neurons or their descending axons to the spinal cord **(Fig. 24-1)** produce weakness through decreased activation of lower motor neurons. In general, distal muscle groups are affected more severely than proximal ones, and axial movements are spared unless the lesion is severe and bilateral. Spasticity is typical but may not be present acutely. Rapid repetitive movements are slowed and coarse, but normal rhythmicity is maintained. With corticobulbar involvement, weakness occurs in the lower face and tongue; extraocular, upper facial, pharyngeal, and jaw muscles are typically spared. Bilateral corticobulbar lesions produce a *pseudobulbar palsy*: dysarthria, dysphagia, dysphonia, and emotional lability accompany bilateral facial weakness and a brisk jaw jerk. Electromyogram (EMG) **(Chap. 446)** shows that with weakness of the upper motor neuron type, motor units have a diminished maximal discharge frequency.

FIGURE 24-1

The corticospinal and bulbospinal upper motor neuron pathways. Upper motor neurons have their cell bodies in layer V of the primary motor cortex (the precentral gyrus, or Brodmann area 4) and in the premotor and supplemental motor cortex (area 6). The upper motor neurons in the primary motor cortex are somatotopically organized (*right side of figure*). Axons of the upper motor neurons descend through the subcortical white matter and the posterior limb of the internal capsule. Axons of the pyramidal or corticospinal system descend through the brainstem in the cerebral peduncle of the midbrain, the basis pontis, and the medullary pyramids. At the cervicomedullary junction, most corticospinal axons decussate into the contralateral corticospinal tract of the lateral spinal cord, but 10–30% remain ipsilateral in the anterior spinal cord. Corticospinal neurons synapse on premotor interneurons, but some—especially in the cervical enlargement and those connecting with motor neurons to distal limb muscles—make direct monosynaptic connections with lower motor neurons. They innervate most densely the lower motor neurons of hand muscles and are involved in the execution of learned, fine movements. Corticobulbar neurons are similar to corticospinal neurons but innervate brainstem motor nuclei. Bulbospinal upper motor neurons influence strength and tone but are not part of the pyramidal system. The descending ventromedial bulbospinal pathways originate in the tectum of the midbrain (tectospinal pathway), the vestibular nuclei (vestibulospinal pathway), and the reticular formation (reticulospinal pathway). These pathways influence axial and proximal muscles and are involved in the maintenance of posture and integrated movements of the limbs and trunk. The descending ventrolateral bulbospinal pathways, which originate predominantly in the red nucleus (rubrospinal pathway), facilitate distal limb muscles. The bulbospinal system sometimes is referred to as the extrapyramidal upper motor neuron system. In all figures, nerve cell bodies and axon terminals are shown, respectively, as *closed circles* and *forks.*

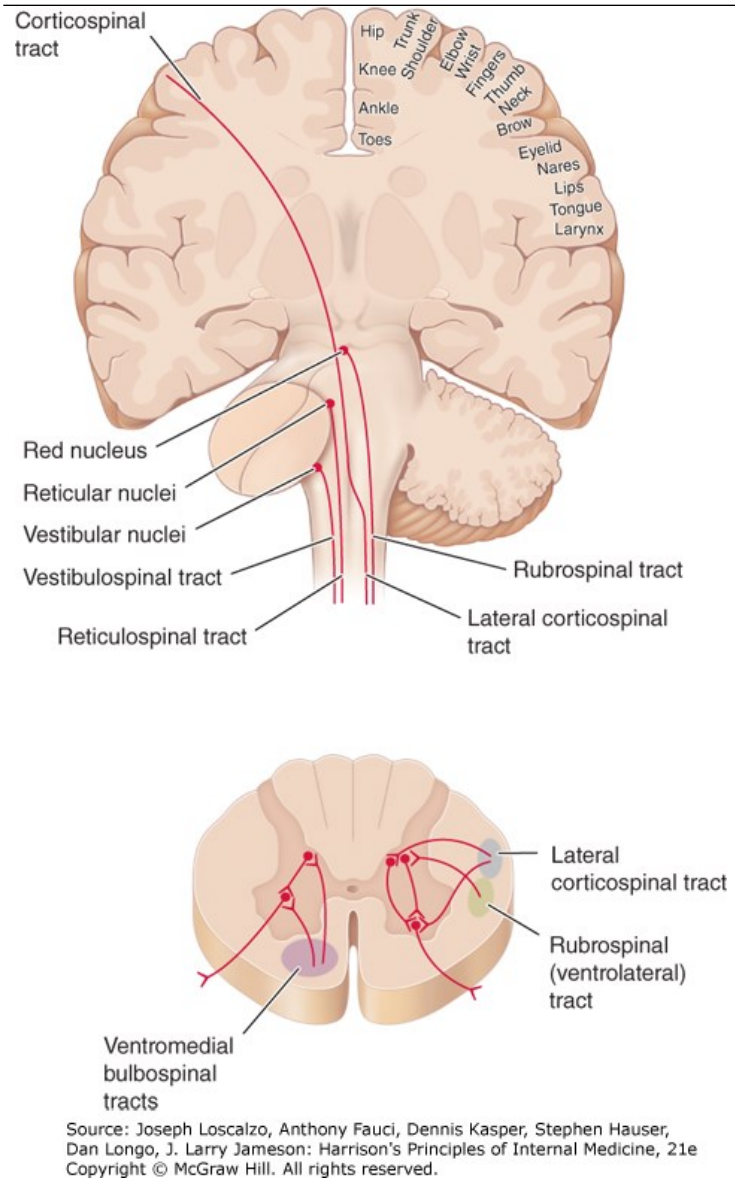
Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 24: Neurologic Causes of Weakness and Paralysis, Stephen L. Hauser

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

system. In all figures, nerve cell bodies and axon terminals are shown, respectively, as *closed circles* and *forks*.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Lower Motor Neuron Weakness

This pattern results from disorders of lower motor neurons in the brainstem motor nuclei and the anterior horn of the spinal cord or from dysfunction of the axons of these neurons as they pass to skeletal muscle **(Fig. 24-2)**. Weakness is due to a decrease in the number of muscle fibers that can be activated through a loss of α motor neurons or disruption of their connections to muscle. Loss of γ motor neurons does not cause weakness but decreases tension on the muscle spindles, which decreases muscle tone and attenuates the stretch reflexes. An absent stretch reflex suggests involvement of spindle afferent fibers.

FIGURE 24-2

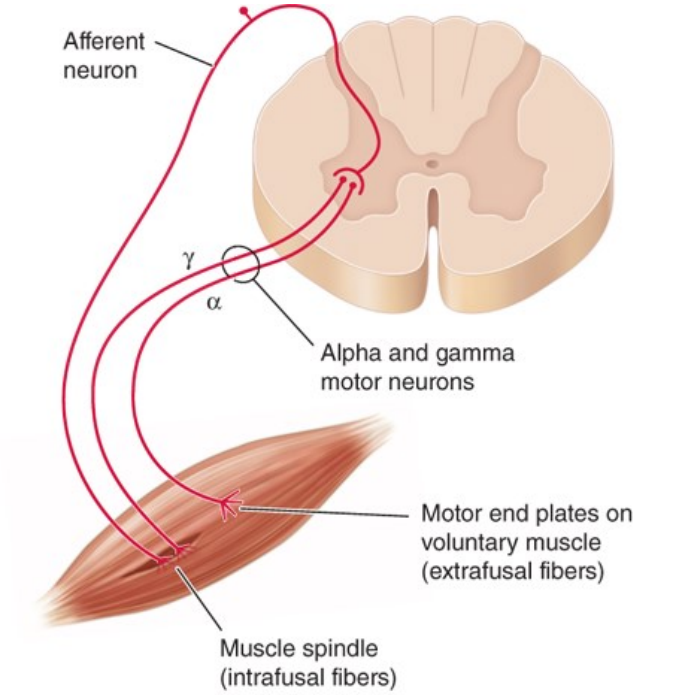
Lower motor neurons are divided into α and γ types. The larger α motor neurons are more numerous and innervate the extrafusal muscle fibers of the motor unit. Loss of α motor neurons or disruption of their axons produces lower motor neuron weakness. The smaller, less numerous γ motor neurons innervate the intrafusal muscle fibers of the muscle spindle and contribute to normal tone and stretch reflexes. The α motor neuron receives direct excitatory input from corticomotoneurons and primary muscle spindle afferents. The α and γ motor neurons also receive excitatory input from other descending upper motor neuron pathways, segmental sensory inputs, and interneurons. The α motor neurons receive direct inhibition from Renshaw cell interneurons, and other interneurons indirectly inhibit the α and γ motor neurons. A muscle stretch (tendon) reflex

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 24: Neurologic Causes of Weakness and Paralysis, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Access Provided by:
SILVERCHAIR

inhibition from Renshaw cell interneurons, and other interneurons indirectly inhibit the α and γ motor neurons. A muscle stretch (tendon) reflex requires the function of all the illustrated structures. A tap on a tendon stretches muscle spindles (which are tonically activated by γ motor neurons) and activates the primary spindle afferent neurons. These neurons stimulate the α motor neurons in the spinal cord, producing a brief muscle contraction, which is the familiar tendon reflex.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

When a motor unit becomes diseased, especially in anterior horn cell diseases, it may discharge spontaneously, producing *fasciculations*. When α motor neurons or their axons degenerate, the denervated muscle fibers also may discharge spontaneously. These single muscle fiber discharges, or *fibrillation potentials,* cannot be seen but can be recorded with EMG. Weakness leads to delayed or reduced recruitment of motor units, with fewer than normal activated at a particular discharge frequency.

Neuromuscular Junction Weakness

Disorders of the neuromuscular junction produce weakness of variable degree and distribution. The number of muscle fibers that are activated varies over time, depending on the state of rest of the neuromuscular junctions. Strength is influenced by preceding activity of the affected muscle. In myasthenia gravis, for example, sustained or repeated contractions of affected muscle decline in strength despite continuing effort **(Chap. 440)**. Thus, fatigable weakness is suggestive of disorders of the neuromuscular junction, which cause functional loss of muscle fibers due to failure of their activation.

Myopathic Weakness

Myopathic weakness is produced by a decrease in the number or contractile force of muscle fibers activated within motor units. With muscular dystrophies, inflammatory myopathies, or myopathies with muscle fiber necrosis, the number of muscle fibers is reduced within many motor units. On EMG, the size of each motor unit action potential is decreased, and motor units must be recruited more rapidly than normal to produce the desired power. Some myopathies produce weakness through loss of contractile force of muscle fibers or through relatively selective involvement of type II (fast) fibers. These myopathies may not affect the size of individual motor unit action potentials and are detected by a discrepancy between the electrical activity and force of a muscle.

Psychogenic Weakness

Weakness may occur without a recognizable organic basis. It tends to be variable, inconsistent, and with a pattern of distribution that cannot be explained on a neuroanatomic basis. On formal testing, antagonists may contract when the patient is supposedly activating the agonist muscle. The

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 24: Neurologic Causes of Weakness and Paralysis, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Hemiparesis

explained on a neuroanatomic basis. On formal testing, antagonists may contract when the patient is supposedly activating the agonist muscle. The severity of weakness is out of keeping with the patient's daily activities.

DISTRIBUTION OF WEAKNESS

Hemiparesis

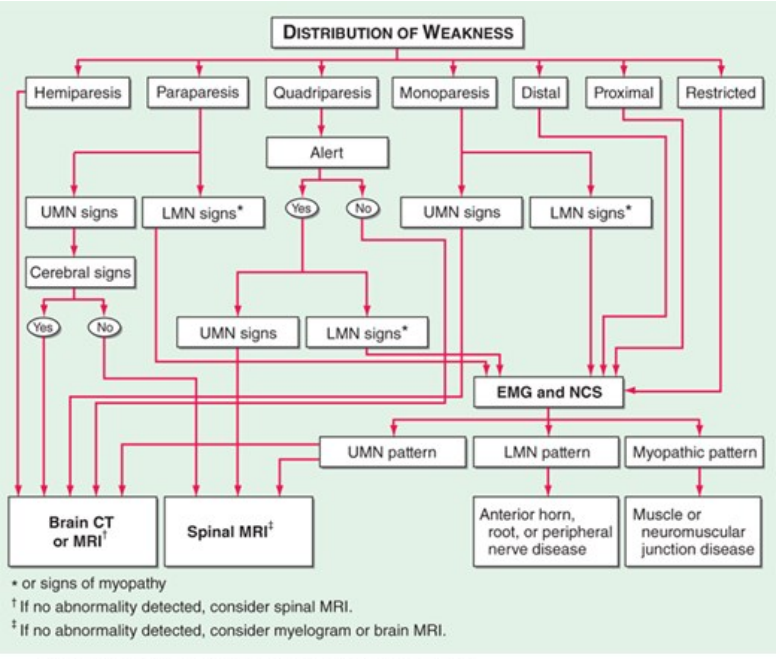
Hemiparesis results from an upper motor neuron lesion above the midcervical spinal cord; most such lesions are above the foramen magnum. The presence of other neurologic deficits helps localize the lesion. Thus language disorders, for example, point to a cortical lesion. Homonymous visual field defects reflect either a cortical or a subcortical hemispheric lesion. A "pure motor" hemiparesis of the face, arm, and leg often is due to a small, discrete lesion in the posterior limb of the internal capsule, cerebral peduncle in the midbrain, or upper pons. Some brainstem lesions produce "crossed paralyses," consisting of ipsilateral cranial nerve signs and contralateral hemiparesis **(Chap. 426)**. The absence of cranial nerve signs or facial weakness suggests that a hemiparesis is due to a lesion in the high cervical spinal cord, especially if associated with Brown-Séquard syndrome, consisting of loss of joint position and vibration sense on the side of the weakness, and loss of pain and temperature sense on the opposite side **(Chap. 442)**.

Acute or episodic hemiparesis usually results from focal structural lesions, particularly vascular etiologies, rapidly expanding lesions, or an inflammatory process. *Subacute hemiparesis* that evolves over days or weeks may relate to subdural hematoma, infectious or inflammatory disorders (e.g., cerebral abscess, fungal granuloma or meningitis, parasitic infection, multiple sclerosis, sarcoidosis), or primary or metastatic neoplasms. AIDS may present with subacute hemiparesis due to toxoplasmosis or primary central nervous system (CNS) lymphoma. *Chronic hemiparesis* that evolves over months usually is due to a neoplasm or vascular malformation, a chronic subdural hematoma, or a degenerative disease.

Investigation of hemiparesis **(Fig. 24-3)** of acute origin usually starts with a CT scan of the brain and laboratory studies. If the CT is normal, or in subacute or chronic cases of hemiparesis, MRI of the brain and/or cervical spine (including the foramen magnum) is performed, depending on the clinical accompaniments.

FIGURE 24-3

An algorithm for the initial workup of a patient with weakness. CT, computed tomography; EMG, electromyography; LMN, lower motor neuron; MRI, magnetic resonance imaging; NCS, nerve conduction studies; UMN, upper motor neuron.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Paraparesis

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

but not touch, position, and vibration sense) suggestive of a central cord syndrome; or hyperreflexia in the legs with normal reflexes in the arms **(Chap. 442)**. Imaging the spinal cord **(Fig. 24-3)** may reveal compressive lesions, infarction (proprioception usually is spared), arteriovenous fistulas

Acute paraparesis is caused most commonly by an intraspinal lesion, but its spinal origin may not be recognized initially if the legs are flaccid and areflexic. Usually, however, there is sensory loss in the legs with an upper level on the trunk; a dissociated sensory loss (loss of pain and temperature but not touch, position, and vibration sense) suggestive of a central cord syndrome; or hyperreflexia in the legs with normal reflexes in the arms **(Chap. 442)**. Imaging the spinal cord **(Fig. 24-3)** may reveal compressive lesions, infarction (proprioception usually is spared), arteriovenous fistulas or other vascular anomalies, or transverse myelitis **(Chap. 442)**.

Diseases of the cerebral hemispheres that produce acute paraparesis include anterior cerebral artery ischemia (shoulder shrug also is affected), superior sagittal sinus or cortical venous thrombosis, and acute hydrocephalus.

Paraparesis may also result from a cauda equina syndrome, for example, after trauma to the low back, a midline disk herniation, or an intraspinal tumor. The sphincters are commonly affected, whereas hip flexion often is spared, as is sensation over the anterolateral thighs. Rarely, paraparesis is caused by a rapidly evolving anterior horn cell disease (such as poliovirus or West Nile virus infection), peripheral neuropathy (such as Guillain-Barré syndrome; **Chap. 447**), or myopathy **(Chap. 449)**.

Subacute or chronic spastic paraparesis is caused by upper motor neuron disease. When associated with lower-limb sensory loss and sphincter involvement, a chronic spinal cord disorder should be considered **(Chap. 442)**. If hemispheric signs are present, a parasagittal meningioma or chronic hydrocephalus is likely. The absence of spasticity in a long-standing paraparesis suggests a lower motor neuron or myopathic etiology.

Investigations typically begin with spinal MRI, but when upper motor neuron signs are associated with drowsiness, confusion, seizures, or other hemispheric signs, brain MRI should also be performed, sometimes as the initial investigation. Electrophysiologic studies are diagnostically helpful when clinical findings suggest an underlying neuromuscular disorder.

Quadriparesis or Generalized Weakness

Generalized weakness may be due to disorders of the CNS or the motor unit. Although the terms often are used interchangeably, *quadriparesis* is commonly used when an upper motor neuron cause is suspected, and *generalized weakness* is used when a disease of the motor units is likely. Weakness from CNS disorders usually is associated with changes in consciousness or cognition and accompanied by spasticity, hyperreflexia, and sensory disturbances. Most neuromuscular causes of generalized weakness are associated with normal mental function, hypotonia, and hypoactive muscle stretch reflexes. The major causes of intermittent weakness are listed in **Table 24-2**. A patient with generalized fatigability without objective weakness may have chronic fatigue syndrome **(Chap. 450)**.

TABLE 24-2

Causes of Episodic Generalized Weakness

1. Electrolyte disturbances, e.g., hypokalemia, hyperkalemia, hypercalcemia, hypernatremia, hyponatremia, hypophosphatemia, hypermagnesemia
2. Muscle disorders
   a. Channelopathies (periodic paralyses)
   b. Metabolic defects of muscle (impaired carbohydrate or fatty acid utilization; abnormal mitochondrial function)
3. Neuromuscular junction disorders
   a. Myasthenia gravis
   b. Lambert-Eaton myasthenic syndrome
4. Central nervous system disorders
   a. Transient ischemic attacks of the brainstem
   b. Transient global cerebral ischemia
   c. Multiple sclerosis
5. Lack of voluntary effort
   a. Anxiety
   b. Pain or discomfort
   c. Somatization disorder

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 24: Neurologic Causes of Weakness and Paralysis, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Quadriparesis with onset over minutes may result from disorders of upper motor neurons (such as from anoxia, hypotension, brainstem or cervical

weakness may have chronic fatigue syndrome (Chap. 450).

TABLE 24-2

Causes of Episodic Generalized Weakness

1. Electrolyte disturbances, e.g., hypokalemia, hyperkalemia, hypercalcemia, hypernatremia, hyponatremia, hypophosphatemia, hypermagnesemia
2. Muscle disorders
   a. Channelopathies (periodic paralyses)
   b. Metabolic defects of muscle (impaired carbohydrate or fatty acid utilization; abnormal mitochondrial function)
3. Neuromuscular junction disorders
   a. Myasthenia gravis
   b. Lambert-Eaton myasthenic syndrome
4. Central nervous system disorders
   a. Transient ischemic attacks of the brainstem
   b. Transient global cerebral ischemia
   c. Multiple sclerosis
5. Lack of voluntary effort
   a. Anxiety
   b. Pain or discomfort
   c. Somatization disorder

ACUTE QUADRIPARESIS

Quadriparesis with onset over minutes may result from disorders of upper motor neurons (such as from anoxia, hypotension, brainstem or cervical cord ischemia, trauma, and systemic metabolic abnormalities) or muscle (electrolyte disturbances, certain inborn errors of muscle energy metabolism, toxins, and periodic paralyses). Onset over hours to weeks may, in addition to these disorders, be due to lower motor neuron disorders such as Guillain-Barré syndrome (Chap. 447).

In obtunded patients, evaluation begins with a CT or MRI scan of the brain. If upper motor neuron signs are present but the patient is alert, the initial test is usually an MRI of the cervical cord. If weakness is lower motor neuron, myopathic, or uncertain in origin, the clinical approach begins with blood studies to determine the level of muscle enzymes and electrolytes and with EMG and nerve conduction studies.

SUBACUTE OR CHRONIC QUADRIPARESIS

Quadriparesis due to upper motor neuron disease may develop over weeks to years from chronic myelopathies, multiple sclerosis, brain or spinal tumors, chronic subdural hematomas, and various metabolic, toxic, and infectious disorders. It may also result from lower motor neuron disease, a chronic neuropathy (in which weakness is often most profound distally), or myopathic weakness (typically proximal).

When quadriparesis develops acutely in obtunded patients, evaluation begins with a CT scan of the brain. If upper motor neuron signs have developed acutely but the patient is alert, the initial test is usually an MRI of the cervical cord. When onset has been gradual, disorders of the cerebral hemispheres, brainstem, and cervical spinal cord can usually be distinguished clinically, and imaging is directed first at the clinically suspected site of pathology. If weakness is lower motor neuron, myopathic, or uncertain in origin, laboratory studies can determine the levels of muscle enzymes and electrolytes, and EMG and nerve conduction studies help to localize the pathologic process (Chap. 449).

Monoparesis

Monoparesis usually is due to lower motor neuron disease, with or without associated sensory involvement. Upper motor neuron weakness occasionally presents as a monoparesis of distal and nonantigravity muscles. Myopathic weakness rarely is limited to one limb.

ACUTE MONOPARESIS

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 24: Neurologic Causes of Weakness and Paralysis, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

motor neuron weakness is likely, evaluation begins with EMG and nerve conduction studies.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

If weakness is predominantly distal and of upper motor neuron type and is not associated with sensory impairment or pain, focal cortical ischemia is likely (Chap. 427); diagnostic possibilities are similar to those for acute hemiparesis. Sensory loss and pain usually accompany acute lower motor neuron weakness; the weakness commonly localizes to a single nerve root or peripheral nerve, but occasionally reflects plexus involvement. If lower motor neuron weakness is likely, evaluation begins with EMG and nerve conduction studies.

SUBACUTE OR CHRONIC MONOPARESIS

Weakness and atrophy that develop over weeks or months are usually of lower motor neuron origin. When associated with sensory symptoms, a peripheral cause (nerve, root, or plexus) is likely; otherwise, anterior horn cell disease should be considered. In either case, an electrodiagnostic study is indicated. If weakness is of the upper motor neuron type, a discrete cortical (precentral gyrus) or cord lesion may be responsible, and appropriate imaging is performed.

Distal Weakness

Involvement of two or more limbs distally suggests lower motor neuron or peripheral nerve disease. Acute distal lower-limb weakness results occasionally from an acute toxic polyneuropathy or cauda equina syndrome. Distal symmetric weakness usually develops over weeks, months, or years and, when associated with numbness, is due to peripheral neuropathy (Chap. 446). Anterior horn cell disease may begin distally but is typically asymmetric and without accompanying numbness (Chap. 437). Rarely, myopathies present with distal weakness (Chap. 449). Electrodiagnostic studies help localize the disorder (Fig. 24-3).

Proximal Weakness

Myopathy often produces symmetric weakness of the pelvic or shoulder girdle muscles (Chap. 449). Diseases of the neuromuscular junction, such as myasthenia gravis (Chap. 448), may present with symmetric proximal weakness often associated with ptosis, diplopia, or bulbar weakness and fluctuate in severity during the day. In anterior horn cell disease, proximal weakness is usually asymmetric, but it may be symmetric especially in genetic forms. Numbness does not occur with any of these diseases. The evaluation usually begins with determination of the serum creatine kinase level and electrophysiologic studies.

Weakness in a Restricted Distribution

Weakness may not fit any of these patterns, being limited, for example, to the extraocular, hemifacial, bulbar, or respiratory muscles. If it is unilateral, restricted weakness usually is due to lower motor neuron or peripheral nerve disease, such as in a facial palsy. Weakness of part of a limb is commonly due to a peripheral nerve lesion such as an entrapment neuropathy. Relatively symmetric weakness of extraocular or bulbar muscles frequently is due to a myopathy (Chap. 449) or neuromuscular junction disorder (Chap. 448). Bilateral facial palsy with areflexia suggests Guillain-Barré syndrome (Chap. 447). Worsening of relatively symmetric weakness with fatigue is characteristic of neuromuscular junction disorders. Asymmetric bulbar weakness usually is due to motor neuron disease. Weakness limited to respiratory muscles is uncommon and usually is due to motor neuron disease, myasthenia gravis, or polymyositis/dermatomyositis (Chap. 365).

ACKNOWLEDGMENT

The editors acknowledge the contributions of Michael J. Aminoff to earlier editions of this chapter.

FURTHER READING

Brazis P et al: *Localization in Clinical Neurology*, 7th ed. Philadelphia, Lippincott William & Wilkins, 2016.

Campbell WW, Barohn RJ: *DeJong's The Neurological Examination*, 8th ed. Philadelphia, Lippincott William & Wilkins, 2019.

Guarantors of Brain: *Aids to the Examination of the Peripheral Nervous System*, 4th ed. Edinburgh, Saunders, 2000.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 24: Neurologic Causes of Weakness and Paralysis, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 25: Numbness, Tingling, and Sensory Loss

Stephen L. Hauser

INTRODUCTION

Normal somatic sensation reflects a continuous monitoring process, little of which reaches consciousness under ordinary conditions. By contrast, disordered sensation, particularly when experienced as painful, is alarming and dominates the patient's attention. Physicians should be able to recognize abnormal sensations by how they are described, know their type and likely site of origin, and understand their implications. **Pain is considered separately in Chap. 13.**

POSITIVE AND NEGATIVE SYMPTOMS

Abnormal sensory symptoms can be divided into two categories: positive and negative. The prototypical positive symptom is tingling (pins and needles); other positive sensory phenomena include itch and altered sensations that are described as pricking, bandlike, lightning-like shooting feelings (lancinations), aching, knifelike, twisting, drawing, pulling, tightening, burning, searing, electrical, or raw feelings. Such symptoms are often painful.

Positive phenomena usually result from trains of impulses generated at sites of lowered threshold or heightened excitability along a peripheral or central sensory pathway. The nature and severity of the abnormal sensation depend on the number, rate, timing, and distribution of ectopic impulses and the type and function of nervous tissue in which they arise. Because positive phenomena represent excessive activity in sensory pathways, they are not necessarily associated with a sensory deficit (loss) on examination.

Negative phenomena represent loss of sensory function and are characterized by diminished or absent feeling that often is experienced as numbness and by abnormal findings on sensory examination. In disorders affecting peripheral sensation, at least one-half of the afferent axons innervating a particular site are probably lost or functionless before a sensory deficit can be demonstrated by clinical examination. If the rate of loss is slow, however, lack of cutaneous feeling may be unnoticed by the patient and difficult to demonstrate on examination, even though few sensory fibers are functioning; if it is rapid, both positive and negative phenomena are usually conspicuous. Subclinical degrees of sensory dysfunction may be revealed by sensory nerve conduction studies or somatosensory-evoked potentials.

Whereas sensory symptoms may be either positive or negative, sensory signs on examination are always a measure of negative phenomena.

TERMINOLOGY

Paresthesias and dysesthesias are general terms used to denote positive sensory symptoms. The term *paresthesias* typically refers to tingling or pins-and-needles sensations but may include a wide variety of other abnormal sensations, except pain; it sometimes implies that the abnormal sensations are perceived spontaneously. The more general term *dysesthesias* denotes all types of abnormal sensations, including painful ones, regardless of whether a stimulus is evident.

Another set of terms refers to sensory abnormalities found on examination. *Hypesthesia* or *hypoesthesia* refers to a reduction of cutaneous sensation to a specific type of testing such as pressure, light touch, and warm or cold stimuli; *anesthesia,* to a complete absence of skin sensation to the same stimuli plus pinprick; and *hypalgesia* or *analgesia*, to reduced or absent pain perception (nociception). *Hyperesthesia* means pain or increased sensitivity in response to touch. Similarly, *allodynia* describes the situation in which a nonpainful stimulus, once perceived, is experienced as painful, even excruciating. An example is elicitation of a painful sensation by application of a vibrating tuning fork. *Hyperalgesia* denotes severe pain in response to a mildly noxious stimulus, and *hyperpathia*, a broad term, encompasses all the phenomena described by hyperesthesia, allodynia, and hyperalgesia. With hyperpathia, the threshold for a sensory stimulus is increased and perception is delayed, but once felt, it is unduly painful.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

absent deep tendon reflexes in the affected limbs. The Romberg sign is positive, which means that the patient sways markedly or topples when asked to stand with feet close together and eyes closed. In severe states of deafferentation involving deep sensation, the patient cannot walk or stand

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Disorders of deep sensation arising from muscle spindles, tendons, and joints affect proprioception (position sense). Manifestations include imbalance (particularly with eyes closed or in the dark), clumsiness of precision movements, and unsteadiness of gait, which are referred to collectively as *sensory ataxia*. Other findings on examination usually, but not invariably, include reduced or absent joint position and vibratory sensibility and absent deep tendon reflexes in the affected limbs. The Romberg sign is positive, which means that the patient sways markedly or topples when asked to stand with feet close together and eyes closed. In severe states of deafferentation involving deep sensation, the patient cannot walk or stand unaided or even sit unsupported. Continuous involuntary movements (*pseudoathetosis*) of the outstretched hands and fingers occur, particularly with eyes closed.

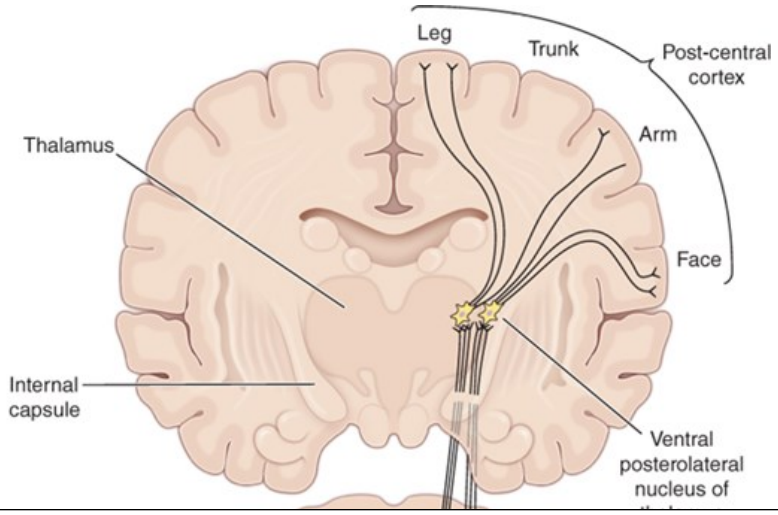
ANATOMY OF SENSATION

Cutaneous receptors are classified by the type of stimulus that optimally excites them. They consist of naked nerve endings (nociceptors, which respond to tissue-damaging stimuli, and thermoreceptors, which respond to noninjurious thermal stimuli) and encapsulated terminals (several types of mechanoreceptor, activated by physical deformation of the skin or stretch of muscles). Each type of receptor has its own set of sensitivities to specific stimuli, size and distinctness of receptive fields, and adaptational qualities.

Afferent peripheral nerve fibers conveying somatosensory information from the limbs and trunk traverse the dorsal roots and enter the dorsal horn of the spinal cord **(Fig. 25-1)**; the cell bodies of first-order neurons are located in the dorsal root ganglia (DRG). In an analogous fashion, sensations from the face and head are conveyed through the trigeminal system (Fig. 441-2). Once fiber tracts enter the spinal cord, the polysynaptic projections of the smaller fibers (unmyelinated and small myelinated), which subserve mainly nociception, itch, temperature sensibility, and touch, cross and ascend in the opposite anterior and lateral columns of the spinal cord, through the brainstem, to the ventral posterolateral (VPL) nucleus of the thalamus and ultimately project to the postcentral gyrus of the parietal cortex and other cortical areas **(Chap. 13)**. This is the *spinothalamic pathway* or *anterolateral system*. The larger fibers, which subserve tactile and position sense and kinesthesia, project rostrally in the posterior and posterolateral columns on the same side of the spinal cord and make their first synapse in the gracile or cuneate nucleus of the lower medulla. Axons of second-order neurons decussate and ascend in the medial lemniscus located medially in the medulla and in the tegmentum of the pons and midbrain and synapse in the VPL nucleus; third-order neurons project to parietal cortex as well as to other cortical areas. This large-fiber system is referred to as the *posterior column–medial lemniscal pathway* (lemniscal, for short). Although the fiber types and functions that make up the spinothalamic and lemniscal systems are relatively well known, many other fibers, particularly those associated with touch, pressure, and position sense, ascend in a diffusely distributed pattern both ipsilaterally and contralaterally in the anterolateral quadrants of the spinal cord. This explains why a complete lesion of the posterior columns of the spinal cord may be associated with little sensory deficit on examination.

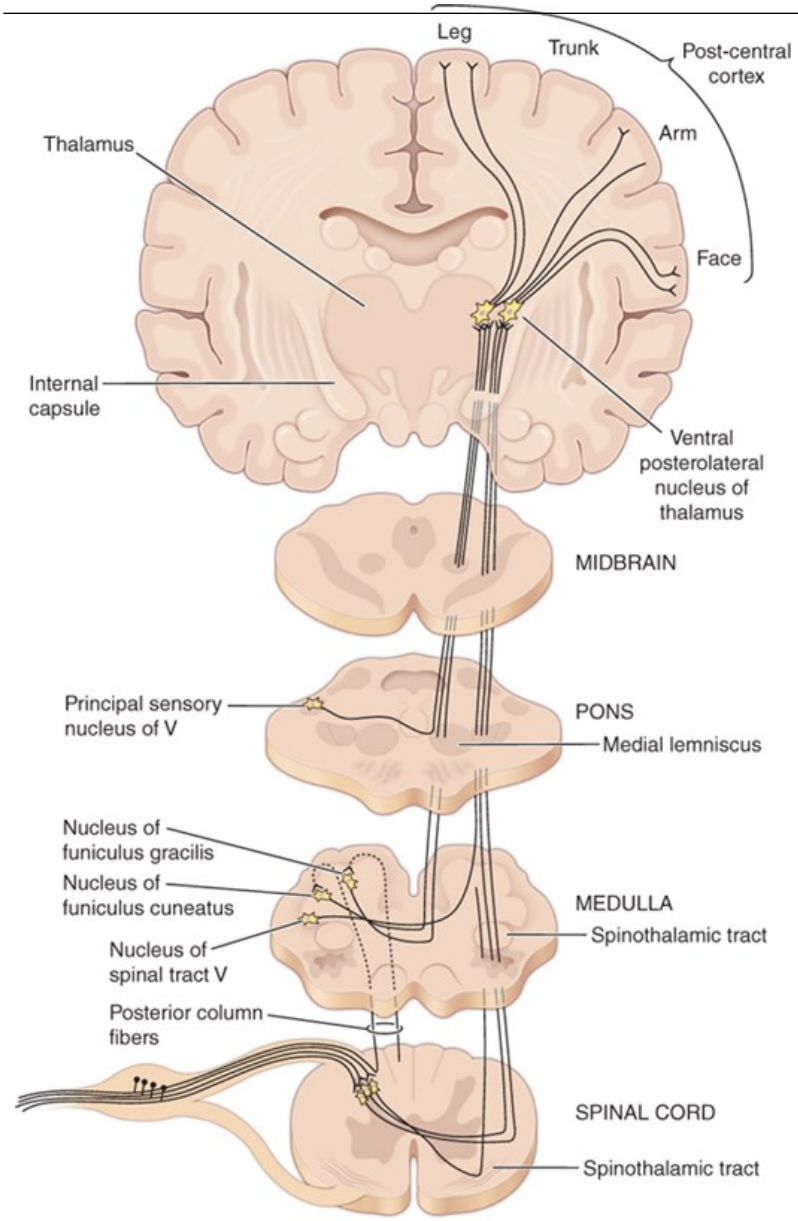
FIGURE 25-1

The main somatosensory pathways. The spinothalamic tract (pain, thermal sense) and the posterior column–lemniscal system (touch, pressure, joint position) are shown. Offshoots from the ascending anterolateral fasciculus (spinothalamic tract) to nuclei in the medulla, pons, and mesencephalon and nuclear terminations of the tract are indicated. *(Reproduced with permission from AH Ropper, MA Samuels: Adams and Victor's Principles of Neurology, 9th ed. New York, McGraw-Hill, 2009.)*



Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

APPROACH TO THE PATIENT

Clinical Examination of Sensation

The main components of the sensory examination are tests of primary sensation (pain, touch, vibration, joint position, and thermal sensation) **(Table 25-1)**. The examiner must depend on patient responses, and this complicates interpretation. Further, examination may be limited in some patients. In a stuporous patient, for example, sensory examination is reduced to observing the briskness of withdrawal in response to a pinch or another noxious stimulus. Comparison of responses on the two sides of the body is essential. In an alert but uncooperative patient, it may not be possible to examine cutaneous sensation, but some idea of proprioceptive function may be gained by noting the patient's best performance of movements requiring balance and precision.

In patients with sensory complaints, testing should begin in the center of the affected region and proceed radially until sensation is perceived as normal. The distribution of any abnormality is defined and compared to root and peripheral nerve territories **(Figs. 25-2 and 25-3)**. Some patients present with sensory symptoms that do not fit an anatomic localization and are accompanied by either no abnormalities or gross inconsistencies on

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

cerebellar systems.

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

normal. The distribution of any abnormality is defined and compared to root and peripheral nerve territories **(Figs. 25-2 and 25-3)**. Some patients present with sensory symptoms that do not fit an anatomic localization and are accompanied by either no abnormalities or gross inconsistencies on examination. The examiner should consider in such cases the possibility of a psychologic cause (see "Psychogenic Symptoms," below). Sensory examination of a patient who has no neurologic complaints can be brief and consist of pinprick, touch, and vibration testing in the hands and feet plus evaluation of stance and gait, including the Romberg maneuver **(Chap. V6)**. Evaluation of stance and gait also tests the integrity of motor and cerebellar systems.

PRIMARY SENSATION

The sense of pain usually is tested with a clean pin, which is then discarded. The patient is asked to close the eyes and focus on the pricking or unpleasant quality of the stimulus, not just the pressure or touch sensation elicited. Areas of hypalgesia should be mapped by proceeding radially from the most hypalgesic site. Temperature sensation to both hot and cold is best tested with small containers filled with water of the desired temperature. An alternative way to test cold sensation is to touch a metal object, such as a tuning fork at room temperature, to the skin. For testing warm temperatures, the tuning fork or another metal object may be held under warm water of the desired temperature and then used. The appreciation of both cold and warmth should be tested because different receptors respond to each. Touch usually is tested with a wisp of cotton, minimizing pressure on the skin. In general, it is better to avoid testing touch on hairy skin because of the profusion of the sensory endings that surround each hair follicle. The patient is tested with the eyes closed and should respond as soon as the stimulus is perceived, indicating its location.

Joint position testing is a measure of proprioception. With the patient's eyes closed, joint position is tested in the distal interphalangeal joint of the great toe and fingers. The digit is held by its sides, distal to the joint being tested, and moved passively while more proximal joints are stabilized—the patient indicates the change in position or direction of movement. If errors are made, more proximal joints are tested. A test of proximal joint position sense, primarily at the shoulder, is performed by asking the patient to bring the two index fingers together with arms extended and eyes closed. Normal individuals can do this accurately, with errors of 1 cm or less.

The sense of vibration is tested with an oscillating tuning fork that vibrates at 128 Hz. Vibration is tested over bony points, beginning distally; in the feet, it is tested over the dorsal surface of the distal phalanx of the big toes and at the malleoli of the ankles, and in the hands, it is tested dorsally at the distal phalanx of the fingers. If abnormalities are found, more proximal sites should be examined. Vibratory thresholds at the same site in the patient and the examiner may be compared for control purposes.

CORTICAL SENSATION

The most commonly used tests of cortical function are two-point discrimination, touch localization, and bilateral simultaneous stimulation, and tests for graphesthesia and stereognosis. Abnormalities of these sensory tests, in the presence of normal primary sensation in an alert cooperative patient, signify a lesion of the parietal cortex or thalamocortical projections. If primary sensation is altered, these cortical discriminative functions usually will be abnormal also. Comparisons should always be made between analogous sites on the two sides of the body because the deficit with a specific parietal lesion is likely to be unilateral.

Two-point discrimination can be tested with calipers, the points of which may be set from 2 mm to several centimeters apart and then applied simultaneously to the test site. On the fingertips, a normal individual can distinguish about a 3-mm separation of points.

Touch localization is performed by light pressure for an instant with the examiner's fingertip or a wisp of cotton wool; the patient, whose eyes are closed, is required to identify the site of touch. *Bilateral simultaneous stimulation* at analogous sites (e.g., the dorsum of both hands) can be carried out to determine whether the perception of touch is extinguished consistently on one side (*extinction* or *neglect*). *Graphesthesia* refers to the capacity to recognize, with eyes closed, letters or numbers drawn by the examiner's fingertip on the palm of the hand. Once again, interside comparison is of prime importance. Inability to recognize numbers or letters is termed *agraphesthesia*.

Stereognosis refers to the ability to identify common objects by palpation, recognizing their shape, texture, and size. Common standard objects such as keys, paper clips, and coins are best used. Patients with normal stereognosis should be able to distinguish a dime from a penny and a nickel from a quarter without looking. Patients should feel the object with only one hand at a time. If they are unable to identify it in one hand, it should be placed in the other for comparison. Individuals who are unable to identify common objects and coins in one hand but can do so in the other are said to have *astereognosis* of the abnormal hand.

QUANTITATIVE SENSORY TESTING

Effective sensory testing devices are commercially available. Quantitative sensory testing is particularly useful for serial evaluation of cutaneous

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Nerve conduction studies and nerve biopsy are important means of investigating the peripheral nervous system, but they do not evaluate the function or structure of cutaneous receptors and free nerve endings or of unmyelinated or thinly myelinated nerve fibers in the nerve trunks. Skin

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

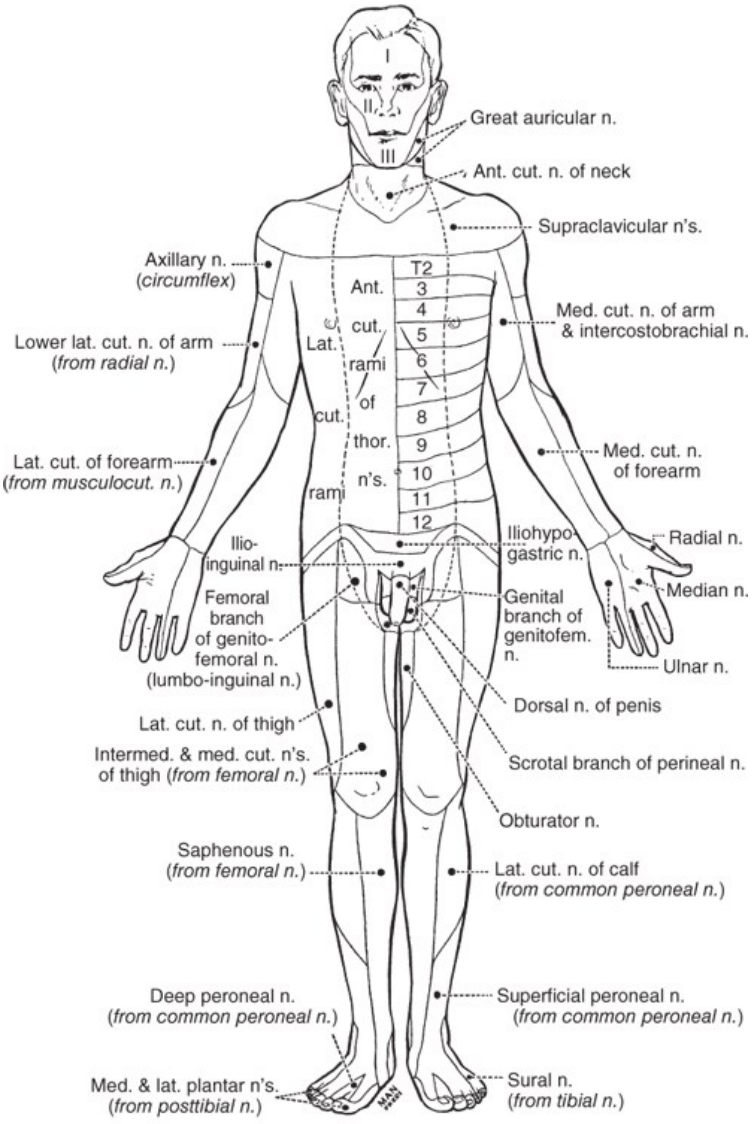
Effective sensory testing devices are commercially available. Quantitative sensory testing is particularly useful for serial evaluation of cutaneous sensation in clinical trials. Threshold testing for touch and vibratory and thermal sensation is the most widely used application.

ELECTRODIAGNOSTIC STUDIES AND NERVE BIOPSY

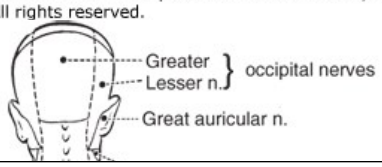
Nerve conduction studies and nerve biopsy are important means of investigating the peripheral nervous system, but they do not evaluate the function or structure of cutaneous receptors and free nerve endings or of unmyelinated or thinly myelinated nerve fibers in the nerve trunks. Skin biopsy can be used to evaluate these structures in the dermis and epidermis.

FIGURE 25-2

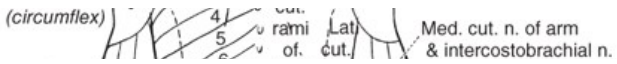
The cutaneous fields of peripheral nerves. *(Reproduced with permission from W Haymaker, B Woodhall: Peripheral Nerve Injuries, 2nd ed. Philadelphia, Saunders, 1953.)*

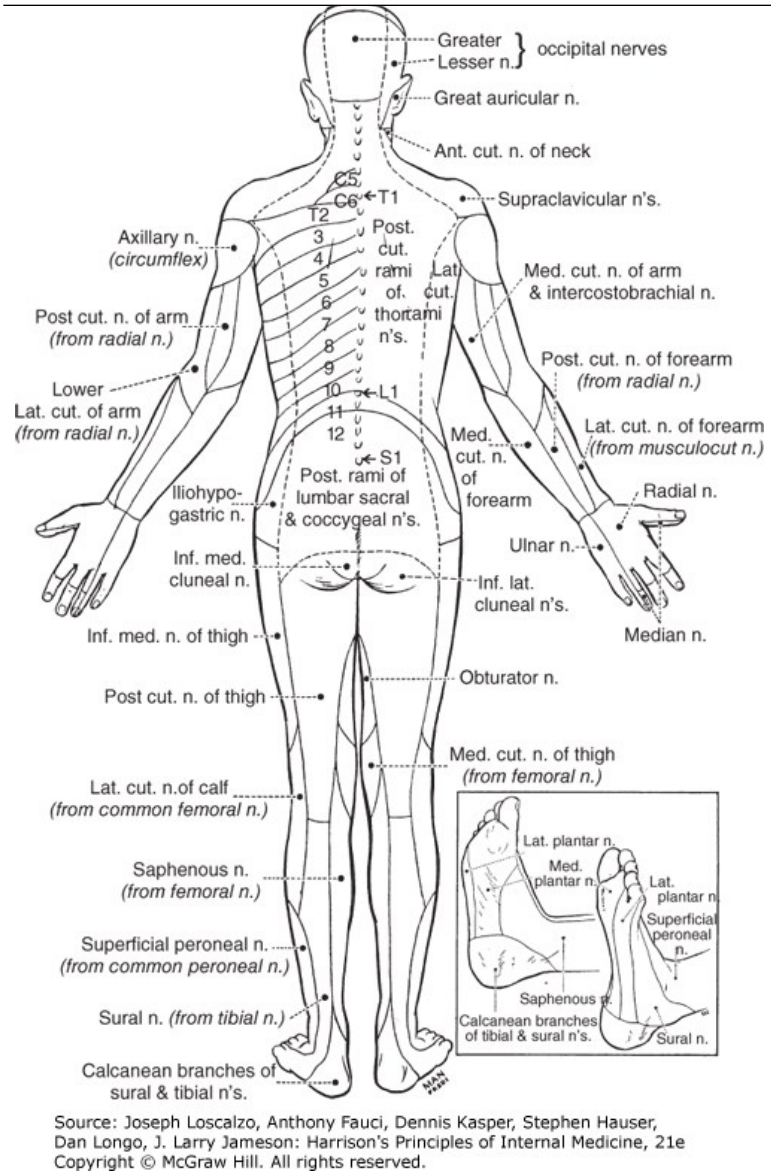


Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.



Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



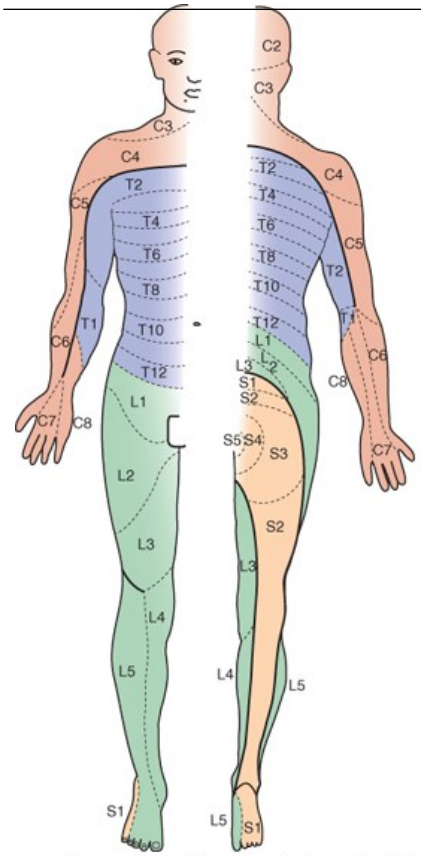


Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 25-3

Distribution of the sensory spinal roots on the surface of the body (dermatomes). *(Reproduced with permission from D Sinclair: Mechanisms of Cutaneous Sensation. Oxford, UK, Oxford University Press, 1981 through PLS Clear.)*

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 25-1

Testing Primary Sensation

SENSE	TEST DEVICE	ENDINGS ACTIVATED	FIBER SIZE MEDIATING	CENTRAL PATHWAY
Pain	Pinprick	Cutaneous nociceptors	Small	SpTh, also D
Temperature, heat	Warm metal object	Cutaneous thermoreceptors for hot	Small	SpTh
Temperature, cold	Cold metal object	Cutaneous thermoreceptors for cold	Small	SpTh
Touch	Cotton wisp, fine brush	Cutaneous mechanoreceptors, also naked endings	Large and small	Lem, also D and SpTh
Vibration	Tuning fork, 128 Hz	Mechanoreceptors, especially pacinian corpuscles	Large	Lem, also D
Joint position	Passive movement of specific joints	Joint capsule and tendon endings, muscle spindles	Large	Lem, also D

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

LOCALIZATION OF SENSORY ABNORMALITIES

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS
Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 25-1

Testing Primary Sensation

SENSE	TEST DEVICE	ENDINGS ACTIVATED	FIBER SIZE MEDIATING	CENTRAL PATHWAY
Pain	Pinprick	Cutaneous nociceptors	Small	SpTh, also D
Temperature, heat	Warm metal object	Cutaneous thermoreceptors for hot	Small	SpTh
Temperature, cold	Cold metal object	Cutaneous thermoreceptors for cold	Small	SpTh
Touch	Cotton wisp, fine brush	Cutaneous mechanoreceptors, also naked endings	Large and small	Lem, also D and SpTh
Vibration	Tuning fork, 128 Hz	Mechanoreceptors, especially pacinian corpuscles	Large	Lem, also D
Joint position	Passive movement of specific joints	Joint capsule and tendon endings, muscle spindles	Large	Lem, also D

Abbreviations: D, diffuse ascending projections in ipsilateral and contralateral anterolateral columns; Lem, posterior column and lemniscal projection, ipsilateral; SpTh, spinothalamic projection, contralateral.

LOCALIZATION OF SENSORY ABNORMALITIES

Sensory symptoms and signs can result from lesions at many different levels of the nervous system from the parietal cortex to the peripheral sensory receptor. Noting their distribution and nature is the most important way to localize their source. Their extent, configuration, symmetry, quality, and severity are the key observations.

Dysesthesias without sensory findings by examination may be difficult to interpret. To illustrate, tingling dysesthesias in an acral distribution (hands and feet) can be systemic in origin, for example, secondary to hyperventilation, or induced by a medication such as acetazolamide. Distal dysesthesias can also be an early event in an evolving polyneuropathy or may herald a myelopathy, such as from vitamin B_{12} deficiency. Sometimes, distal dysesthesias have no definable basis. In contrast, dysesthesias that correspond in distribution to that of a particular peripheral nerve structure denote a lesion at that site. For instance, dysesthesias restricted to the fifth digit and the adjacent one-half of the fourth finger on one hand reliably point to disorder of the ulnar nerve, most commonly at the elbow.

Nerve and Root

In focal nerve trunk lesions, sensory abnormalities are readily mapped and generally have discrete boundaries (Figs. 25-2 and 25-3). Root ("radicular") lesions frequently are accompanied by deep, aching pain along the course of the related nerve trunk. With compression of a fifth lumbar (L5) or first sacral (S1) root, as from a ruptured intervertebral disk, sciatica (radicular pain relating to the sciatic nerve trunk) is a common manifestation **(Chap. 17)**. With a lesion affecting a single root, sensory deficits may be minimal or absent because adjacent root territories overlap extensively.

Isolated mononeuropathies may cause symptoms beyond the territory supplied by the affected nerve, but abnormalities on examination typically are confined to expected anatomic boundaries. In multiple mononeuropathies, symptoms and signs occur in discrete territories supplied by different individual nerves and—as more nerves are affected—may simulate a polyneuropathy if deficits become confluent. With polyneuropathies, sensory deficits are generally graded, distal, and symmetric in distribution **(Chap. 446)**. Dysesthesias followed by numbness begin in the toes and ascend

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

brainstem, but an upper level of the sensory disturbance may then be found on the trunk and other evidence of a central lesion may be present, such as sphincter involvement or signs of an upper motor neuron lesion **(Chap. 24)**. Although most polyneuropathies are pansensory and affect all

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

individual nerves and—as more nerves are affected—may simulate a polyneuropathy if deficits become confluent. With polyneuropathies, sensory deficits are generally graded, distal, and symmetric in distribution **(Chap. 446)**. Dysesthesias, followed by numbness, begin in the toes and ascend symmetrically. When dysesthesias reach the knees, they usually also have appeared in the fingertips. The process is nerve length–dependent, and the deficit is often described as "stocking glove" in type. Involvement of both hands and feet also occurs with lesions of the upper cervical cord or the brainstem, but an upper level of the sensory disturbance may then be found on the trunk and other evidence of a central lesion may be present, such as sphincter involvement or signs of an upper motor neuron lesion **(Chap. 24)**. Although most polyneuropathies are pansensory and affect all modalities of sensation, selective sensory dysfunction according to nerve fiber size may occur. Small-fiber polyneuropathies are characterized by burning, painful dysesthesias with reduced pinprick and thermal sensation but with sparing of proprioception, motor function, and deep tendon reflexes. Touch is involved variably; when it is spared, the sensory pattern is referred to as exhibiting *sensory dissociation*. Sensory dissociation may occur also with spinal cord lesions **(Chap. 442)**. Large-fiber polyneuropathies are characterized by vibration and position sense deficits, imbalance, absent tendon reflexes, and variable motor dysfunction but preservation of most cutaneous sensation. Dysesthesias, if present at all, tend to be tingling or bandlike in quality.

Sensory neuronopathy (or ganglionopathy) is characterized by widespread but asymmetric sensory loss occurring in a non-length-dependent manner so that it may occur proximally or distally, and in the arms, legs, or both. Pain and numbness progress to sensory ataxia and impairment of all sensory modalities over time. This condition is usually paraneoplastic or idiopathic in origin **(Chaps. 94 and 445)** or related to an autoimmune disease, particularly Sjögren's syndrome **(Chap. 361)**.

Spinal Cord

(See also Chap. 442) If the spinal cord is transected, all sensation is lost below the level of transection. Bladder and bowel function also are lost, as is motor function. Lateral hemisection of the spinal cord produces the Brown-Séquard syndrome, with absent pain and temperature sensation contralaterally and loss of proprioceptive sensation and power ipsilaterally below the lesion (see Figs. 25-1 and **442-1**); ipsilateral pain or hyperesthesia may also occur.

Numbness or paresthesias in both feet may arise from a spinal cord lesion; this is especially likely when the upper level of the sensory loss extends to the trunk. When all extremities are affected, the lesion is probably in the cervical region or brainstem unless a peripheral neuropathy is responsible. The presence of upper motor neuron signs **(Chap. 24)** supports a central lesion; a hyperesthetic band on the trunk may suggest the level of involvement.

A dissociated sensory loss can reflect spinothalamic tract involvement in the spinal cord, especially if the deficit is unilateral and has an upper level on the torso. Bilateral spinothalamic tract involvement occurs with lesions affecting the center of the spinal cord, such as in syringomyelia. There is a dissociated sensory loss with impairment of pinprick and temperature appreciation but relative preservation of light touch, position sense, and vibration appreciation.

Dysfunction of the posterior columns in the spinal cord or of the posterior root entry zone may lead to a bandlike sensation around the trunk or a feeling of tight pressure in one or more limbs. Flexion of the neck sometimes leads to an electric shock–like sensation that radiates down the back and into the legs (Lhermitte's sign) in patients with a cervical lesion affecting the posterior columns, such as from multiple sclerosis, cervical spondylosis, or following irradiation to the cervical region.

Brainstem

Crossed patterns of sensory disturbance, in which one side of the face and the opposite side of the body are affected, localize to the lateral medulla. Here a small lesion may damage both the ipsilateral descending trigeminal tract and the ascending spinothalamic fibers subserving the opposite arm, leg, and hemitorso (see "Lateral medullary syndrome" in **Fig. 426-7**). A lesion in the tegmentum of the pons and midbrain, where the lemniscal and spinothalamic tracts merge, causes pansensory loss contralaterally.

Thalamus

Hemisensory disturbance with tingling numbness from head to foot is often thalamic in origin but also can arise from the anterior parietal region. If abrupt in onset, the lesion is likely to be due to a small stroke (lacunar infarction), particularly if localized to the thalamus. Occasionally, with lesions affecting the VPL nucleus or adjacent white matter, a syndrome of thalamic pain, also called *Déjerine-Roussy syndrome*, may ensue. The persistent, unrelenting unilateral pain often is described in dramatic terms.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

With lesions of the parietal lobe involving either the cortex or subjacent white matter, the most prominent symptoms are contralateral hemineglect, hemi-inattention, and a tendency not to use the affected hand and arm. On cortical sensory testing (e.g., two-point discrimination, graphesthesia),

unrelenting unilateral pain often is described in dramatic terms.

Cortex

With lesions of the parietal lobe involving either the cortex or subjacent white matter, the most prominent symptoms are contralateral hemineglect, hemi-inattention, and a tendency not to use the affected hand and arm. On cortical sensory testing (e.g., two-point discrimination, graphesthesia), abnormalities are often found but primary sensation is usually intact. Anterior parietal infarction may present as a pseudothalamic syndrome with contralateral loss of primary sensation from head to toe. Dysesthesias or a sense of numbness and, rarely, a painful state may also occur.

Focal Sensory Seizures

These seizures generally are due to lesions in the area of the postcentral or precentral gyrus. The principal symptom of focal sensory seizures is tingling, but additional, more complex sensations may occur, such as a rushing feeling, a sense of warmth, or a sense of movement without detectable motion. Symptoms typically are unilateral; commonly begin in the arm or hand, face, or foot; and often spread in a manner that reflects the cortical representation of different bodily parts, as in a Jacksonian march. Their duration is variable; seizures may be transient, lasting only for seconds, or persist for an hour or more. Focal motor features may supervene, often becoming generalized with loss of consciousness and tonic-clonic jerking.

Psychogenic Symptoms

Sensory symptoms may have a psychogenic basis. Such symptoms may be generalized or have an anatomic boundary that is difficult to explain neurologically, for example, circumferentially at the groin or shoulder or around a specific joint. Pain is common, but the nature and intensity of any sensory disturbances are variable. The diagnosis should not be one of exclusion but based on suggestive findings that are otherwise difficult to explain, such as midline splitting of impaired vibration, pinprick, or light touch appreciation; variability or poor reproducibility of sensory deficits; or normal performance of tasks requiring sensory input that is seemingly abnormal on formal testing, such as good performance with eyes closed of the finger-to-nose test despite an apparent loss of position sense in the upper limb. The side with abnormal sensation may be confused when the limbs are placed in an unusual position, such as crossed behind the back. Sensory complaints should not be regarded as psychogenic simply because they are unusual.

TREATMENT

Management is based on treatment of the underlying condition. Symptomatic treatment of acute and chronic pain is discussed in **Chap. 13**. Dysesthesias, when severe and persistent, may respond to anticonvulsants (carbamazepine, 100–1000 mg/d; gabapentin, 300–3600 mg/d; or pregabalin, 50–300 mg/d), antidepressants (amitriptyline, 25–150 mg/d; nortriptyline, 25–150 mg/d; desipramine, 100–300 mg/d; or venlafaxine, 75–225 mg/d).

ACKNOWLEDGMENTS

The editors acknowledge the contributions of Michael J. Aminoff to earlier editions of this chapter.

FURTHER READING

Brazis P et al: *Localization in Clinical Neurology*, 7th ed. Philadelphia, Lippincott William & Wilkins, 2016.

Campbell WW, Barohn RJ: *DeJong's the Neurologic Examination*, 8th ed. Philadelphia, Wolters Kluwer, 2020.

Waxman S: *Clinical Neuroanatomy*, 29th ed. New York, McGraw Hill Education, 2020.

Downloaded 2022-1-3 5:45 P Your IP is 141.216.78.40
Chapter 25: Numbness, Tingling, and Sensory Loss, Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 26: Gait Disorders, Imbalance, and Falls

Jessica M. Baker

PREVALENCE, MORBIDITY, AND MORTALITY

Gait and balance problems are common in the elderly and contribute to the risk of falls and injury. Gait disorders have been described in 15% of individuals aged >65. By age 80, one person in four will use a mechanical aid to assist with ambulation. Among those aged ≥85, the prevalence of gait abnormality approaches 40%. In epidemiologic studies, gait disorders are consistently identified as a major risk factor for falls and injury.

ANATOMY AND PHYSIOLOGY

An upright bipedal gait depends on the successful integration of postural control and locomotion. These functions are widely distributed in the central nervous system. The biomechanics of bipedal walking are complex, and the performance is easily compromised by a neurologic deficit at any level. Command and control centers in the brainstem, cerebellum, and forebrain modify the action of spinal pattern generators to promote stepping. While a form of "fictive locomotion" can be elicited from quadrupedal animals after spinal transection, this capacity is limited in primates. Step generation in primates is dependent on locomotor centers in the pontine tegmentum, midbrain, and subthalamic region. Locomotor synergies are executed through the reticular formation and descending pathways in the ventromedial spinal cord. Cerebral control provides a goal and purpose for walking and is involved in avoidance of obstacles and adaptation of locomotor programs to context and terrain.

Postural control requires the maintenance of the center of mass over the base of support through the gait cycle. Unconscious postural adjustments maintain standing balance: long latency responses are measurable in the leg muscles, beginning 110 milliseconds after a perturbation. Forward motion of the center of mass provides propulsive force for stepping, but failure to maintain the center of mass within stability limits results in falls. The anatomic substrate for dynamic balance has not been well defined, but the vestibular nucleus and midline cerebellum contribute to balance control in animals. Patients with damage to these structures have impaired balance while standing and walking.

Standing balance depends on good-quality sensory information about the position of the body center with respect to the environment, support surface, and gravitational forces. Sensory information for postural control is primarily generated by the visual system, the vestibular system, and proprioceptive receptors in the muscle spindles and joints. A healthy redundancy of sensory afferent information is generally available, but loss of two of the three pathways is sufficient to compromise standing balance. Balance disorders in older individuals sometimes result from multiple insults in the peripheral sensory systems (e.g., visual loss, vestibular deficit, peripheral neuropathy) that critically degrade the quality of afferent information needed for balance stability.

Older patients with cognitive impairment appear to be particularly prone to falls and injury. There is a growing body of literature on the use of attentional resources to manage gait and balance. Walking is generally considered to be unconscious and automatic, but the ability to walk while attending to a cognitive task (*dual-task walking*) may be compromised in the elderly. Older patients with deficits in executive function may have particular difficulty in managing the attentional resources needed for dynamic balance when distracted.

DISORDERS OF GAIT

Disorders of gait may be attributed to neurologic and nonneurologic causes, although significant overlap often exists. The *antalgic gait* results from avoidance of pain associated with weight bearing and is commonly seen in osteoarthritis. Asymmetry is a common feature of gait disorders due to contractures and other orthopedic deformities. Impaired vision rounds out the list of common nonneurologic causes of gait disorders.

Neurologic gait disorders are disabling and equally important to address. The heterogeneity of gait disorders observed in clinical practice reflects the large network of neural systems involved in the task. Walking is vulnerable to neurologic disease at every level. Gait disorders have been classified

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

by the adaptive response. Some common patterns of abnormal gait are summarized next. Gait disorders can also be classified by etiology **(Table 26-1)**.

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

large network of neural systems involved in the task. Walking is vulnerable to neurologic disease at every level. Gait disorders have been classified descriptively on the basis of abnormal physiology and biomechanics. One problem with this approach is that many failing gaits look fundamentally similar. This overlap reflects common patterns of adaptation to threatened balance stability and declining performance. *The gait disorder observed clinically must be viewed as the product of a neurologic deficit and a functional adaptation.* Unique features of the failing gait are often overwhelmed by the adaptive response. Some common patterns of abnormal gait are summarized next. Gait disorders can also be classified by etiology **(Table 26-1)**.

TABLE 26-1

Prevalence of Neurologic Gait Disorders

NEUROLOGIC GAIT DISORDER	NO. (%)[a]	TOTAL NUMBER[b]	CAUSES (NO.)
Single neurologic gait disorder	81 (69%)		
Sensory ataxic	22 (18%)	46	Peripheral sensory neuropathy (46)
Parkinsonian	19 (16%)	34	Parkinson's disease (18), drug-induced parkinsonism (8), dementia with parkinsonism (4), parkinsonism (4)
Higher level	9 (8%)	31	Vascular encephalopathy (20), normal pressure hydrocephalus (1), severe dementia (7), hypoxic ischemic encephalopathy (1), unknown (1)
Cerebellar ataxic	7 (6%)	10	Cerebellar stroke (3), cerebellar lesion due to multiple sclerosis (1), severe essential tremor (3), postvaccinal cerebellitis (1), chronic alcohol abuse (1), multiple system atrophy (1)
Cautious	7 (6%)	7	Idiopathic, associated fear of falling (7)
Paretic/hypotonic	6 (5%)	14	Neurogenic claudication (7), diabetic neuropathy (1), nerve lesion due to trauma or surgery (4), distal paraparesis after Guillain-Barré syndrome (1), unknown (2)
Spastic	6 (5%)	7	Ischemic stroke (3), intracerebral hemorrhage (3), congenital (1)
Vestibular ataxic	4 (3%)	6	Bilateral vestibulopathy (3), recent vestribular neuronitis (1), recent Ménière's attack (1), acoustic neuroma with surgery (1)
Dyskinetic	1 (1%)	4	Levodopa-induced dyskinesia (3), chorea (1)
Multiple neurologic gait disorders	36 (30%)		
Total	117		

[a]Percentage of individuals with a single gait disorder. [b]Includes individuals with multiple gait disorders.

Note: Of 117 patients with a neurologic gait disorder, 81 had a single neurologic gait disorder; the remainder (36) had multiple neurologic gait disorders.

Source: Reproduced with modifications from P Mahlknecht et al: PLoS One 8:e69627, 2013.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

a slippery surface. Arms are often held abducted. This disorder is both common and nonspecific. It is, in essence, an adaptation to a perceived postural threat. There may be an associated fear of falling. This disorder can be observed in more than one-third of older patients with gait impairment.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

CAUTIOUS GAIT

The term *cautious gait* is used to describe the patient who walks with an abbreviated stride, widened base, and lowered center of mass, as if walking on a slippery surface. Arms are often held abducted. This disorder is both common and nonspecific. It is, in essence, an adaptation to a perceived postural threat. There may be an associated fear of falling. This disorder can be observed in more than one-third of older patients with gait impairment. Physical therapy often improves walking to the degree that follow-up observation may reveal a more specific underlying disorder.

STIFF-LEGGED GAIT

Spastic gait is characterized by stiffness in the legs, an imbalance of muscle tone, and a tendency to circumduct and scuff the feet. The disorder reflects compromise of corticospinal command and overactivity of spinal reflexes. The patient may walk on the toes. In extreme instances, the legs cross due to increased tone in the adductors ("scissoring" gait). Upper motor neuron signs are present on physical examination. The disorder may be cerebral or spinal in origin.

Myelopathy from cervical spondylosis is a common cause of spastic or spastic-ataxic gait in the elderly. Demyelinating disease and trauma are the leading causes of myelopathy in younger patients. In chronic progressive myelopathy of unknown cause, a workup with laboratory and imaging tests may establish a diagnosis. A structural lesion, such as a tumor or a spinal vascular malformation, should be excluded with appropriate testing. **Spinal cord disorders are discussed in detail in Chap. 442.**

With cerebral spasticity, asymmetry is common, the upper extremities are usually involved, and dysarthria is often an associated feature. Common causes include vascular disease (stroke), multiple sclerosis, motor neuron disease, and perinatal nervous system injury (cerebral palsy).

Other stiff-legged gaits include dystonia **(Chap. 436)** and stiff-person syndrome **(Chap. 94)**. Dystonia is a disorder characterized by sustained muscle contractions resulting in repetitive twisting movements and abnormal posture. It often has a genetic basis. Dystonic spasms can produce plantar flexion and inversion of the feet, sometimes with torsion of the trunk. In autoimmune stiff-person syndrome, exaggerated lordosis of the lumbar spine and overactivation of antagonist muscles restrict trunk and lower-limb movement and result in a wooden or fixed posture.

PARKINSONISM, FREEZING GAIT, AND OTHER MOVEMENT DISORDERSS

Parkinson's disease **(Chap. 435)** is common, affecting 1% of the population >65 years of age. The stooped posture, shuffling gait, and decreased arm swing are characteristic and distinctive features. Patients sometimes accelerate (festinate) with walking, display retropulsion, or exhibit a tendency to turn en bloc. The step-to-step variability of the parkinsonian gait also contributes to falls, which are a major source of morbidity, particularly later in the disease course. Dopamine replacement improves step length, arm swing, turning speed, and gait initiation. There is increasing evidence that deficits in cholinergic circuits in the pedunculopontine nucleus and cortex contribute to the gait disorder of Parkinson's disease. Cholinesterase inhibitors such as donepezil and rivastigmine have been shown in early studies to significantly decrease gait variability, instability, and fall frequency, even in the absence of cognitive impairment, perhaps through improvement in attention.

Freezing is defined as a brief, episodic absence of forward progression of the feet, despite the intention to walk. Freezing may be triggered by approaching a narrow doorway or crowd, may be overcome by visual cueing, and contributes to fall risk. Gait freezing is present in approximately one-quarter of Parkinson's patients within 5 years of onset, and its frequency increases further over time. In treated patients, end-of-dose gait freezing is a common problem that may improve with more frequent administration of dopaminergic drugs or with use of monoamine oxidase type B inhibitors such as rasagiline or selegiline **(Chap. 435)**.

Freezing of gait is also common in other neurodegenerative disorders associated with parkinsonism, including progressive supranuclear palsy (PSP), multiple-system atrophy, and corticobasal degeneration. Patients with these disorders frequently present with axial stiffness, postural instability, and a shuffling, freezing gait while lacking the characteristic pill-rolling tremor of Parkinson's disease. The gait of PSP is typically more erect compared with the stooped posture of typical Parkinson's disease, and falls within the first year also suggest the possibility of PSP. The gait of vascular parkinsonism tends to be broad-based and shuffling with reduced arm swing bilaterally; disproportionate involvement of gait early in the disease course differentiates this entity from Parkinson's disease.

Hyperkinetic movement disorders also produce characteristic and recognizable disturbances in gait. In Huntington's disease **(Chap. 436)**, the unpredictable occurrence of choreic movements gives the gait a dancing quality. Tardive dyskinesia is the cause of many odd, stereotypic gait disorders seen in patients chronically exposed to antipsychotics and other drugs that block the D$_2$ dopamine receptor. *Orthostatic tremor* is a high

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FRONTAL GAIT DISORDER

ACCESS Medicine
Access Provided by: SILVERCHAIR

unpredictable occurrence of choreic movements gives the gait a dancing quality. Tardive dyskinesia is the cause of many odd, stereotypic gait disorders seen in patients chronically exposed to antipsychotics and other drugs that block the D_2 dopamine receptor. *Orthostatic tremor* is a high-frequency, low-amplitude tremor predominantly involving the lower extremities. Patients often report shakiness or unsteadiness on standing and improvement with sitting or walking. Falls are common. The tremor is often only appreciable by palpating the legs while standing.

FRONTAL GAIT DISORDER

Frontal gait disorder, also known as higher-level gait disorder, is common in the elderly and has a variety of causes. The term is used to describe a shuffling, freezing gait with imbalance, and other signs of higher cerebral dysfunction. Typical features include a wide base of support, a short stride, shuffling along the floor, and difficulty with starts and turns. Many patients exhibit a difficulty with gait initiation that is descriptively characterized as the "slipping clutch" syndrome or gait ignition failure. The term *lower-body parkinsonism* is also used to describe such patients. Strength is generally preserved, and patients are able to make stepping movements when not standing and maintaining their balance at the same time. This disorder is best considered a higher-level motor control disorder, as opposed to an apraxia **(Chap. 30)**, though the term *gait apraxia* persists in the literature.

The most common cause of frontal gait disorder is vascular disease, particularly subcortical small-vessel disease in the deep frontal white matter and centrum ovale. Over three-quarters of patients with subcortical vascular dementia demonstrate gait abnormalities; decreased arm swing and a stooped posture are particularly prevalent features. The clinical syndrome also includes dysarthria, pseudobulbar affect (emotional disinhibition), increased tone, and hyperreflexia in the lower limbs.

Normal pressure (communicating) hydrocephalus (NPH) in adults also presents with a similar gait disorder **(Chap. 431)**. Other features of the diagnostic triad (mental changes, incontinence) may be absent in a substantial number of patients. MRI demonstrates ventricular enlargement, an enlarged flow void about the aqueduct, periventricular white matter change, and high-convexity tightness (disproportionate widening of the sylvian fissures versus the cortical sulci). A lumbar puncture or dynamic test is necessary to confirm a diagnosis of NPH. Neurodegenerative dementias and mass lesions of the frontal lobes cause a similar clinical picture and can be differentiated from vascular disease and hydrocephalus by neuroimaging.

CEREBELLAR GAIT ATAXIA

Disorders of the cerebellum **(Chap. 439)** have a dramatic impact on gait and balance. Cerebellar gait ataxia is characterized by a wide base of support, lateral instability of the trunk, erratic foot placement, and decompensation of balance when attempting to walk on a narrow base. Difficulty maintaining balance when turning is often an early feature. Patients are unable to walk tandem heel to toe and display truncal sway in narrow-based or tandem stance. They show considerable variation in their tendency to fall in daily life.

Causes of cerebellar ataxia in older patients include stroke, trauma, tumor, and neurodegenerative disease such as multiple-system atrophy **(Chap. 440)** and various forms of hereditary cerebellar degeneration **(Chap. 439)**. A short expansion at the site of the fragile X mutation (*fragile X premutation*) has been associated with gait ataxia in older men. Alcohol causes an acute and chronic cerebellar ataxia. In patients with ataxia due to cerebellar degeneration, MRI demonstrates the extent and topography of cerebellar atrophy.

SENSORY ATAXIA

As reviewed earlier in this chapter, balance depends on high-quality afferent information from the visual and the vestibular systems and proprioception. When this information is lost or degraded, balance during locomotion is impaired and instability results. The sensory ataxia of tabetic neurosyphilis is a classic example. The contemporary equivalent is the patient with neuropathy affecting large fibers. Vitamin B_{12} deficiency is a treatable cause of large-fiber sensory loss in the spinal cord and peripheral nervous system. Joint position and vibration sense are diminished in the lower limbs. The stance in such patients is destabilized by eye closure; they often look down at their feet when walking and do poorly in the dark. **Table 26-2** compares sensory ataxia with cerebellar ataxia and frontal gait disorder.

TABLE 26-2

Features of Cerebellar Ataxia, Sensory Ataxia, and Frontal Gait Disorders

FEATURE	CEREBELLAR ATAXIA	SENSORY ATAXIA	FRONTAL GAIT
Base of support	Wide-based	Narrow base, looks down	Wide-based
Stride	Irregular, lurching	Regular with path deviation	Short, shuffling

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

Table 26-2 compares sensory ataxia with cerebellar ataxia and frontal gait disorder.

TABLE 26-2

Features of Cerebellar Ataxia, Sensory Ataxia, and Frontal Gait Disorders

FEATURE	CEREBELLAR ATAXIA	SENSORY ATAXIA	FRONTAL GAIT
Base of support	Wide-based	Narrow base, looks down	Wide-based
Velocity	Variable	Slow	Very slow
Stride	Irregular, lurching	Regular with path deviation	Short, shuffling
Romberg test	+/–	Unsteady, falls	+/–
Heel → shin	Abnormal	+/–	Normal
Initiation	Normal	Normal	Hesitant
Turns	Unsteady	+/–	Hesitant, multistep
Postural instability	+	+++	++++ Poor postural synergies rising from a chair
Falls	Late event	Frequent	Frequent

NEUROMUSCULAR DISEASE

Patients with neuromuscular disease often have an abnormal gait, occasionally as a presenting feature. With distal weakness (peripheral neuropathy), the step height is increased to compensate for foot drop, and the sole of the foot may slap on the floor during weight acceptance, termed the *steppage gait*. Patients with myopathy or muscular dystrophy more typically exhibit proximal weakness. Weakness of the hip girdle may result in some degree of excess pelvic sway during locomotion. The stooped posture of lumbar spinal stenosis ameliorates pain from the compression of the cauda equina occurring with a more upright posture while walking and may mimic early parkinsonism.

TOXIC AND METABOLIC DISORDERS

Chronic toxicity from medications and metabolic disturbances can impair motor function and gait. Examination may reveal mental status changes, asterixis, or myoclonus. Static equilibrium is disturbed, and such patients are easily thrown off balance. Disequilibrium is particularly evident in patients with chronic renal disease and those with hepatic failure, in whom asterixis may impair postural support. Sedative drugs, especially neuroleptics and long-acting benzodiazepines, affect postural control and increase the risk for falls. These disorders are especially important to recognize because they are often treatable.

FUNCTIONAL GAIT DISORDER

Functional neurologic disorders (formerly “psychogenic”) are common in practice, and the presentation often involves gait. Sudden onset, inconsistent deficits, waxing and waning course, incongruence of symptoms with an organic lesion, and improvement with distraction are key features. Phenomenology is variable; extreme slow motion, an inappropriately overcautious gait, odd gyrations of posture with wastage of muscular energy, astasia–abasia (inability to stand and walk), bouncing, and foot stiffness (dystonia) have been described. Falls are rare, and there are often discrepancies between examination findings and the patient’s functional status. Preceding stress or trauma is variably present, and its absence does not preclude the diagnosis of a functional gait disorder. Functional gait disorders may be challenging to diagnose and should be differentiated from the slowness and psychomotor retardation seen in certain patients with major depression.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

When reviewing the history, it is helpful to inquire about the onset and progression of disability. Initial awareness of an unsteady gait often follows a fall. Stepwise evolution or sudden progression suggests vascular disease. Gait disorder may be associated with urinary urgency and incontinence

Access Provided by:
SILVERCHAIR

APPROACH TO THE PATIENT

Slowly Progressive Disorder of Gait

When reviewing the history, it is helpful to inquire about the onset and progression of disability. Initial awareness of an unsteady gait often follows a fall. Stepwise evolution or sudden progression suggests vascular disease. Gait disorder may be associated with urinary urgency and incontinence, particularly in patients with cervical spine disease or hydrocephalus. It is always important to review the use of alcohol and medications that affect gait and balance. Information on localization derived from the neurologic examination can be helpful in narrowing the list of possible diagnoses.

Gait observation provides an immediate sense of the patient's degree of disability. Arthritic and antalgic gaits are recognized by observation, although neurologic and orthopedic problems may coexist. Characteristic patterns of abnormality are sometimes seen, although, as stated previously, failing gaits often look fundamentally similar. Cadence (steps per minute), velocity, and stride length can be recorded by timing a patient over a fixed distance. Watching the patient rise from a chair provides a good functional assessment of balance.

Brain imaging studies may be informative in patients with an undiagnosed disorder of gait. MRI is sensitive for cerebral lesions of vascular or demyelinating disease and is a good screening test for occult hydrocephalus. Patients with recurrent falls are at risk for subdural hematoma. As mentioned earlier, many elderly patients with gait and balance difficulty have white matter abnormalities in the periventricular region and centrum semiovale. While these lesions may be an incidental finding, a substantial burden of white matter disease will ultimately impact cerebral control of locomotion.

DISORDERS OF BALANCE

DEFINITION, ETIOLOGY, AND MANIFESTATIONS

Balance is the ability to maintain equilibrium—a dynamic state in which one's center of mass is controlled with respect to the lower extremities, gravity, and the support surface despite external perturbations. The reflexes required to maintain upright posture require input from cerebellar, vestibular, and somatosensory systems; the premotor cortex and corticospinal and reticulospinal tracts mediate output to axial and proximal limb muscles. These responses are physiologically complex, and the anatomic representation they entail is not well understood. Failure can occur at any level and presents as difficulty maintaining posture while standing and walking.

The history and physical examination may differentiate underlying causes of imbalance. Patients with *cerebellar* ataxia do not generally complain of dizziness, although balance is visibly impaired. Neurologic examination reveals a variety of cerebellar signs. Postural compensation may prevent falls early on, but falls are inevitable with disease progression. The progression of neurodegenerative ataxia is often measured by the number of years to loss of stable ambulation.

Vestibular disorders **(Chap. 22)** have symptoms and signs that fall into three categories: (1) vertigo (the subjective inappropriate perception or illusion of movement); (2) nystagmus (involuntary eye movements); and (3) impaired standing balance. Not every patient has all manifestations. Patients with vestibular deficits related to ototoxic drugs may lack vertigo or obvious nystagmus, but their balance is impaired on standing and walking, and they cannot navigate in the dark. Laboratory testing is available to investigate vestibular deficits.

Somatosensory deficits also produce imbalance and falls. There is often a subjective sense of insecure balance and fear of falling. Postural control is compromised by eye closure (*Romberg's sign*); these patients also have difficulty navigating in the dark. A dramatic example is provided by the patient with autoimmune subacute sensory neuropathy, which is sometimes a paraneoplastic disorder **(Chap. 94)**. Compensatory strategies enable such patients to walk in the virtual absence of proprioception, but the task requires active visual monitoring.

Patients with *higher-level disorders of equilibrium* have difficulty maintaining balance in daily life and may present with falls. Their awareness of balance impairment may be reduced. Patients taking sedating medications are in this category.

FALLS

Falls are common in the elderly; over one-third of people aged >65 who are living in the community fall each year. This number is even higher in

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

proportion of spinal cord injuries due to falls in individuals aged >65 years has doubled in the past decade, perhaps due to increasing activity in this age group. Some falls result in a prolonged time lying on the ground; fractures and CNS injury are a particular concern in this context.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Falls are common in the elderly; over one-third of people aged >65 who are living in the community fall each year. This number is even higher in nursing homes and hospitals. Elderly people are not only at higher risk for falls but are also more likely to suffer serious complications due to medical comorbidities such as osteoporosis. Hip fractures result in hospitalization, can lead to nursing home admission, and are associated with an increased mortality risk in the subsequent year. Falls may result in brain or spinal injury, the history of which may be difficult for the patient to provide. The proportion of spinal cord injuries due to falls in individuals aged >65 years has doubled in the past decade, perhaps due to increasing activity in this age group. Some falls result in a prolonged time lying on the ground; fractures and CNS injury are a particular concern in this context.

For each person who is physically disabled, there are others whose functional independence is limited by anxiety and fear of falling. Nearly one in five elderly individuals voluntarily restricts his or her activity because of fear of falling. With loss of ambulation, the quality of life diminishes, and rates of morbidity and mortality increase.

RISK FACTORS FOR FALLS

Risk factors for falls may be *intrinsic* (e.g., gait and balance disorders) or *extrinsic* (e.g., polypharmacy, environmental factors); some risk factors are modifiable. The presence of multiple risk factors is associated with a substantially increased risk of falls. **Table 26-3** summarizes a meta-analysis of studies establishing the principal risk factors for falls. Polypharmacy (use of four or more prescription medications) has also been identified as an important risk factor.

TABLE 26-3

Meta-Analysis of Risk Factors for Falls in Older Persons

RISK FACTOR	MEAN RR (OR)	RANGE
Muscle weakness	4.4	1.5–10.3
History of falls	3.0	1.7–7.0
Gait deficit	2.9	1.3–5.6
Balance deficit	2.9	1.6–5.4
Use assistive device	2.6	1.2–4.6
Visual deficit	2.5	1.6–3.5
Arthritis	2.4	1.9–2.9
Impaired ADL	2.3	1.5–3.1
Depression	2.2	1.7–2.5
Cognitive impairment	1.8	1.0–2.3
Age >80 years	1.7	1.1–2.5

Abbreviations: ADL, activity of daily living; OR, odds ratio from retrospective studies; RR, relative risk from prospective studies.

Source: Reproduced with permission from Guideline for the Prevention of Falls in Older Persons. J Am Geriatr Soc 49:664, 2001.

ASSESSMENT OF THE PATIENT WITH FALLS

The most productive approach is to identify the high-risk patient prospectively, before there is a serious injury. All community-dwelling adults should

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

History

Access Provided by:
SILVERCHAIR

The most productive approach is to identify the high-risk patient prospectively, before there is a serious injury. All community-dwelling adults should be asked annually about falls and whether or not fear of falling limits daily activities. The Timed Up and Go ("TUG") test involves timing a patient as they stand up from a chair, walk 10 feet, turn, and then sit down. Patients with a history of falls or those requiring >12 s to complete the TUG test are at high risk for falls and should undergo further assessment.

History

The history surrounding a fall is often problematic or incomplete, and the underlying mechanism or cause may be difficult to establish in retrospect. Patients should be queried about any provoking factors (including head turn, standing) or prodromal symptoms, such as dizziness, vertigo, presyncopal symptoms, or focal weakness. A history of baseline mobility and medical comorbidities should be elicited. Patients at particular risk include those with mental status changes or dementia. Medications should be reviewed, with particular attention to benzodiazepines, opioids, antipsychotics, antiepileptics, antidepressants, antiarrhythmics, and diuretics, all of which are associated with an increased risk of falls. It is equally important to distinguish *mechanical falls* (those caused by tripping or slipping) due to purely extrinsic or environmental factors from those in which a modifiable intrinsic factor contributes. *Recurrent falls* may indicate an underlying gait or balance disorder. Falls associated with loss of consciousness (syncope, seizure) may require appropriate cardiac or neurologic evaluation and intervention **(Chaps. 21 and 425)**, although a patient's report of change in consciousness may be unreliable.

Physical Examination

Examination of the patient with falls should include a basic cardiac examination, including orthostatic blood pressure if indicated by history, and observation of any orthopedic abnormalities. Mental status is easily assessed while obtaining a history from the patient; the remainder of the neurologic examination should include visual acuity, strength and sensation in the lower extremities, muscle tone, and cerebellar function, with particular attention to gait and balance as described earlier in this chapter.

Fall Patterns

The description of a fall event may provide further clues to the underlying etiology. While there is no standard nosology of falls, some common clinical patterns may emerge and provide a clue.

DROP ATTACKS AND COLLAPSING FALLS

Drop attacks and collapsing falls are associated with a sudden loss of postural tone. Patients may report that their legs just "gave out" underneath them or that they "collapsed in a heap." Syncope or orthostatic hypotension may be a factor in some such falls. Neurologic causes are relatively rare but include atonic seizures, myoclonus, and intermittent obstruction of the foramen of Monro by a colloid cyst of the third ventricle causing acute obstructive hydrocephalus. An emotional trigger suggests cataplexy. While collapsing falls are more common among older patients with vascular risk factors, drop attacks should not be confused with vertebrobasilar ischemic attacks.

TOPPLING FALLS

Some patients maintain tone in antigravity muscles but fall over like a tree trunk, as if postural defenses had disengaged. Causes include cerebellar pathology and lesions of the vestibular system. There may be a consistent direction to such falls. Toppling falls are an early feature of progressive supranuclear palsy, and a late feature of Parkinson's disease, once postural instability has developed. Thalamic lesions causing truncal instability (*thalamic astasia*) may also contribute to this type of fall.

FALLS DUE TO GAIT FREEZING

Freezing of gait is seen in Parkinson's disease and related disorders. The feet stick to the floor and the center of mass keeps moving, resulting in a disequilibrium from which the patient has difficulty recovering, resulting in a forward fall. Similarly, patients with Parkinson's disease and festinating gait may find their feet unable to keep up and may thus fall forward.

FALLS RELATED TO SENSORY LOSS

Patients with somatosensory, visual, or vestibular deficits are prone to falls. These patients have particular difficulty dealing with poor illumination or

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FALLS RELATED TO WEAKNESS

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Patients with somatosensory, visual, or vestibular deficits are prone to falls. These patients have particular difficulty dealing with poor illumination or walking on uneven ground. They often report subjective imbalance, apprehension, and fear of falling. These patients may be especially responsive to a rehabilitation-based intervention.

FALLS RELATED TO WEAKNESS

Patients who lack strength in antigravity muscles have difficulty rising from a chair or maintaining their balance after a perturbation. These patients are often unable to get up after a fall and may have to remain on the floor for a prolonged period until help arrives. If due to deconditioning, this is often treatable. Resistance strength training can increase muscle mass and leg strength, even for people in their eighties and nineties.

TREATMENT

Interventions to Reduce the Risk of Falls and Injury

Efforts should be made to define the mechanism underlying falls in a given patient, as specific treatment may be possible once a diagnosis is established. Orthostatic changes in blood pressure and pulse should be recorded. Medications (including over-the-counter) should be reviewed, reevaluating benefits and burdens of medications that might increase fall risk. Treatment of cataracts and avoidance of multifocal lenses could be considered for patients whose falls result from vision impairment. A home visit to look for environmental hazards can be helpful. A variety of modifications may be recommended to improve safety, including improved lighting, installation of grab bars and nonslip surfaces, and use of adaptive equipment.

Home- and group-based exercise programs focusing on leg strength and balance, physical therapy, and use of assistive devices reduce fall risk in individuals with a history of falls or disorders of gait and balance. Rehabilitative interventions aim to improve muscle strength and balance stability and to make the patient more resistant to injury. High-intensity resistance strength training with weights and machines is useful to improve muscle mass, even in frail older patients. Improvements realized in posture and gait should translate to reduced risk of falls and injury. Sensory balance training is another approach to improving balance stability. Measurable gains can be made in a few weeks of training, and benefits can be maintained over 6 months by a 10- to 20-min home exercise program. This strategy is particularly successful in patients with vestibular and somatosensory balance disorders. The National Institute on Aging provides online examples of balance exercises for older adults. A Tai Chi exercise program has been demonstrated to reduce the risk of falls and injury in patients with Parkinson's disease. Cognitive training, including dual-task training, may improve mobility in older adults with cognitive impairment.

ACKNOWLEDGEMENTS

I am grateful to Dr. Lewis R. Sudarsky for his substantial contributions to earlier versions of this chapter.

FURTHER READING

American Geriatrics Society, British Geriatrics Society, American Academy of Orthopedic Surgeons Panel on Falls Prevention: Guideline for the prevention of falls in older persons. J Am Geriatr Soc 49:664, 2001. [PubMed: 11380764]

Ganz D, Latham N: Prevention of falls in community-dwelling older adults. N Engl J Med 382:734, 2020. [PubMed: 32074420]

National Institute on Aging: Exercise and Physical Activity. Available from https://www-nia-nih-gov.libproxy.umflint.edu/health/exercise-physical-activity. Accessed April 25, 2021.

Nutt JG: Classification of gait and balance disorders. Adv Neurol 87:135, 2001. [PubMed: 11347216]

Pirker W, Katzenschlager R: Gait disorders in adults and the elderly. Wien Klin Wochenschr 129:81, 2017. [PubMed: 27770207]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 26: Gait Disorders, Imbalance, and Falls, Jessica M. Baker
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 27: Confusion and Delirium

S. Andrew Josephson; Bruce L. Miller

INTRODUCTION

Confusion, a mental and behavioral state of reduced comprehension, coherence, and capacity to reason, is one of the most common problems encountered in medicine, accounting for a large number of emergency department visits, hospital admissions, and inpatient consultations. *Delirium*, a term used to describe an acute confusional state, remains a major cause of morbidity and mortality, costing billions of dollars yearly in health care costs in the United States alone. Despite increased efforts targeting awareness of this condition, delirium often goes unrecognized in the face of evidence that it is usually the cognitive manifestation of serious underlying medical or neurologic illness.

CLINICAL FEATURES OF DELIRIUM

A multitude of terms are used to describe patients with delirium, including *encephalopathy, acute brain failure, acute confusional state*, and *postoperative* or *intensive care unit (ICU) psychosis*. Delirium has many clinical manifestations, but it is defined as a relatively acute decline in cognition that fluctuates over hours or days. The hallmark of delirium is a deficit of attention, although all cognitive domains—including memory, executive function, visuospatial tasks, and language—are variably involved. Associated symptoms that may be present in some cases include altered sleep-wake cycles, perceptual disturbances such as hallucinations or delusions, affect changes, and autonomic findings that include heart rate and blood pressure instability.

Delirium is a clinical diagnosis that is made only at the bedside. Two subtypes have been described—hyperactive and hypoactive—based on differential psychomotor features. The cognitive syndrome associated with severe alcohol withdrawal (i.e., “delirium tremens”) remains the classic example of the hyperactive subtype, featuring prominent hallucinations, agitation, and hyperarousal, often accompanied by life-threatening autonomic instability. In striking contrast is the hypoactive subtype, exemplified by benzodiazepine intoxication, in which patients are withdrawn and quiet, with prominent apathy and psychomotor slowing.

This dichotomy between subtypes of delirium is a useful construct, but patients often fall somewhere along a spectrum between the hyperactive and hypoactive extremes, sometimes fluctuating from one to the other. Therefore, clinicians must recognize this broad range of presentations of delirium to identify all patients with this potentially reversible cognitive disturbance. Hyperactive patients are often easily recognized by their characteristic severe agitation, tremor, hallucinations, and autonomic instability. Patients who are quietly hypoactive are more often overlooked on the medical wards and in the ICU.

The reversibility of delirium is emphasized because many etiologies, such as infection and medication effects, can be treated easily. The long-term cognitive consequences of delirium remain an area of active research. Some episodes of delirium continue for weeks, months, or even years. The persistence of delirium in some patients and its high recurrence rate may be due to inadequate initial treatment of the underlying etiology. In other instances, delirium appears to cause permanent neuronal damage and long-term cognitive decline. Therefore, prevention strategies are important to implement. Even if an episode of delirium completely resolves, there may be lingering effects of the disorder; a patient’s recall of events after delirium varies widely, ranging from complete amnesia to repeated reexperiencing of the frightening period of confusion, similar to what is seen in patients with posttraumatic stress disorder.

RISK FACTORS

An effective primary prevention strategy for delirium begins with identification of high-risk patients. Some hospital systems have initiated comprehensive delirium programs that screen most or all patients upon admission or before elective surgery; positive screens trigger a host of focused prevention measures. Multiple validated scoring systems have been developed as a screen for asymptomatic patients, many of which emphasize well-established risk factors for delirium

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

are truly independent risk factors is uncertain. Other predisposing factors include sensory deprivation, such as preexisting hearing and visual impairment, as well as indices for poor overall health, including baseline immobility, malnutrition, and underlying medical or neurologic illness.

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

focused prevention measures. Multiple validated scoring systems have been developed as a screen for asymptomatic patients, many of which emphasize well-established risk factors for delirium.

The two most consistently identified risk factors are older age and baseline cognitive dysfunction. Individuals who are aged >65 or exhibit low scores on standardized tests of cognition develop delirium upon hospitalization at a rate approaching 50%. Whether age and baseline cognitive dysfunction are truly independent risk factors is uncertain. Other predisposing factors include sensory deprivation, such as preexisting hearing and visual impairment, as well as indices for poor overall health, including baseline immobility, malnutrition, and underlying medical or neurologic illness.

In-hospital risks for delirium include the use of bladder catheterization, physical restraints, sleep and sensory deprivation, and the addition of three or more new medications. Avoiding such risks remains a key component of delirium prevention as well as treatment. Surgical and anesthetic risk factors for the development of postoperative delirium include procedures such as those involving cardiopulmonary bypass, inadequate or excessive treatment of pain in the immediate postoperative period, and perhaps specific agents such as inhalational anesthetics.

The relationship between delirium and dementia **(Chap. 29)** is complicated by significant overlap between the two conditions, and it is not always simple to distinguish between them. Dementia and preexisting cognitive dysfunction serve as major risk factors for delirium, and at least two-thirds of cases of delirium occur in patients with coexisting underlying dementia. A form of dementia with parkinsonism, *dementia with Lewy bodies* **(Chap. 434)**, is characterized by a fluctuating course, prominent visual hallucinations, parkinsonism, and an attentional deficit that clinically resembles hyperactive delirium; patients with this condition are particularly vulnerable to delirium. Delirium in the elderly often reflects an insult to a brain that is vulnerable due to an underlying neurodegenerative condition. Therefore, the development of delirium sometimes heralds the onset of a previously unrecognized brain disorder, and after the acute delirious episode has cleared, careful screening for an underlying condition should occur in the outpatient setting.

EPIDEMIOLOGY

Delirium is common, but its reported incidence has varied widely with the criteria used to define this disorder. Estimates of delirium in hospitalized patients range from 10% to >50%, with higher rates reported for elderly patients and patients undergoing hip surgery. Older patients in the ICU have especially high rates of delirium that approach 75%. The condition is not recognized in up to one-third of delirious inpatients, and the diagnosis is especially problematic in the ICU environment, where cognitive dysfunction is often difficult to appreciate in the setting of serious systemic illness and sedation. Delirium in the ICU should be viewed as an important manifestation of organ dysfunction not unlike liver, kidney, or heart failure. Outside the acute hospital setting, delirium occurs in nearly one-quarter of patients in nursing homes and in 50–80% of those at the end of life. These estimates emphasize the remarkably high frequency of this cognitive syndrome in older patients, a population that continues to grow.

An episode of delirium was previously viewed as a transient condition that carried a benign prognosis. It is now recognized as a disorder with substantial morbidity and mortality, and that often represents the first manifestation of a serious underlying illness. Estimates of in-hospital mortality rates among delirious patients range from 25% to 33%, similar to mortality rates due to sepsis. Patients with an in-hospital episode of delirium have a fivefold higher mortality rate in the months after their illness compared with age matched nondelirious hospitalized patients. Delirious hospitalized patients also have a longer length of stay, are more likely to be discharged to a nursing home, have a higher frequency of readmission, and are more likely to experience subsequent episodes of delirium and cognitive decline; as a result, this condition has an enormous economic cost.

PATHOGENESIS

The pathogenesis and anatomy of delirium are incompletely understood. The attentional deficit that serves as the neuropsychological hallmark of delirium has a diffuse localization within the brainstem, thalamus, prefrontal cortex, and parietal lobes. Rarely, focal lesions such as ischemic strokes have led to delirium in otherwise healthy persons; right parietal and medial dorsal thalamic lesions have been reported most commonly, pointing to the importance of these areas in delirium pathogenesis. In most cases, however, delirium results from widespread disturbances in cortical and subcortical regions of the brain. Electroencephalogram (EEG) usually reveals symmetric slowing, a nonspecific finding that supports diffuse cerebral dysfunction.

Multiple neurotransmitter abnormalities, proinflammatory factors, and specific genes likely play a role in the pathogenesis of delirium. Deficiency of acetylcholine may play a key role, and medications with anticholinergic properties can commonly precipitate delirium. As noted earlier, patients with preexisting dementia are particularly susceptible to episodes of delirium. Alzheimer's disease **(Chap. 431)**, dementia with Lewy bodies **(Chap. 434)**, and Parkinson's disease dementia **(Chap. 435)** are all associated with cholinergic deficiency due to degeneration of acetylcholine-producing neurons in the basal forebrain. In addition, other neurotransmitters are also likely to be involved in this diffuse cerebral disorder. For example, increases in

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

on a healthy young adult but produce a florid delirium in an elderly person with known underlying dementia, although even healthy young persons

Multiple neurotransmitter abnormalities, proinflammatory factors, and specific genes likely play a role in the pathogenesis of delirium. Deficiency of in the basal forebrain. In addition, other neurotransmitters are also likely to be involved in this diffuse cerebral disorder. For example, increases in dopamine can lead to delirium, and patients with Parkinson's disease treated with dopaminergic medications can develop a delirium-like state that features visual hallucinations, fluctuations, and confusion.

Not all individuals exposed to the same insult will develop signs of delirium. A low dose of an anticholinergic medication may have no cognitive effects on a healthy young adult but produce a florid delirium in an elderly person with known underlying dementia, although even healthy young persons develop delirium with very high doses of anticholinergic medications. This concept of delirium developing as the result of an insult in predisposed individuals is currently the most widely accepted pathogenic construct. Therefore, if a previously healthy individual with no known history of cognitive illness develops delirium in the setting of a relatively minor insult such as elective surgery or hospitalization, an unrecognized underlying neurologic illness such as a neurodegenerative disease, multiple previous strokes, or another diffuse cerebral cause should be considered. In this context, delirium can be viewed as a "stress test for the brain" whereby exposure to known inciting factors such as systemic infection and offending drugs can unmask a decreased cerebral reserve and herald a serious underlying and potentially treatable illness. New blood-based biomarkers for specific dementias may soon be available to help predict people at risk for delirium before surgical procedures or hospitalization.

APPROACH TO THE PATIENT

Delirium

Because the diagnosis of delirium is clinical and is made at the bedside, a careful history and physical examination are necessary in evaluating patients with possible confusional states. Screening tools can aid physicians and nurses in identifying patients with delirium, including the Confusion Assessment Method (CAM); the Nursing Delirium Screening Scale (NuDESC); the Organic Brain Syndrome Scale; the Delirium Rating Scale; and, in the ICU, the ICU version of the CAM and the Delirium Detection Score. Using the well-validated CAM, a diagnosis of delirium is made if there is (1) an acute onset and fluctuating course and (2) inattention accompanied by either (3) disorganized thinking or (4) an altered level of consciousness **(Table 27-1)**. These scales may not identify the full spectrum of patients with delirium, and all patients who are acutely confused should be presumed delirious regardless of their presentation due to the wide variety of possible clinical features. A course that fluctuates over hours or days and may worsen at night (termed *sundowning*) is typical but not essential for the diagnosis. Observation will usually reveal an altered level of consciousness or a deficit of attention. Other features that are sometimes present include alteration of sleep-wake cycles, thought disturbances such as hallucinations or delusions, autonomic instability, and changes in affect.

HISTORY

It may be difficult to elicit an accurate history in delirious patients who have altered levels of consciousness or impaired attention. Information from a collateral source such as a spouse or another family member is therefore invaluable. The three most important pieces of history are the patient's baseline cognitive function, the time course of the present illness, and current medications.

Premorbid cognitive function can be assessed through the collateral source or, if needed, via a review of outpatient records. Delirium by definition represents a change that is relatively acute and usually developing over hours to days, from a cognitive baseline. An acute confusional state is nearly impossible to diagnose without some knowledge of baseline cognitive function. Without this information, many patients with dementia or longstanding depression may be mistaken as delirious during a single initial evaluation. Patients with a more hypoactive, apathetic presentation with psychomotor slowing may be identified as being different from baseline only through conversations with family members. A number of validated instruments have been shown to diagnose cognitive dysfunction accurately using a collateral source, including the modified Blessed Dementia Rating Scale and the Clinical Dementia Rating (CDR). Baseline cognitive impairment is common in patients with delirium. Even when no such history of cognitive impairment is elicited, there should still be a high suspicion for a previously unrecognized underlying neurologic disorder.

Establishing the time course of cognitive change is important not only to make a diagnosis of delirium but also to correlate the onset of the illness with potentially treatable etiologies such as recent medication changes or symptoms of systemic infection.

Medications remain a common cause of delirium, especially compounds with anticholinergic or sedative properties. It is estimated that nearly one-third of all cases of delirium are secondary to medications, especially in the elderly. Medication histories should include all prescription as well as over-the-counter and herbal substances taken by the patient and any recent changes in dosing or formulation, including substitution of generics for brand-name medications.

Other important elements of the history include screening for symptoms of organ failure or systemic infection, which often contributes to delirium in the elderly. A history of illicit drug use, alcoholism, or toxin exposure is common in younger delirious patients. Finally, asking the patient and

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The general physical examination in a delirious patient should include careful screening for signs of infection such as fever, tachypnea, pulmonary consolidation, heart murmur, and meningismus. The patient's fluid status should be assessed; both dehydration and fluid overload with resultant

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

in the elderly. A history of illicit drug use, alcoholism, or toxin exposure is common in younger delirious patients. Finally, asking the patient and collateral source about other symptoms that may accompany delirium, such as depression, may help identify potential therapeutic targets.

PHYSICAL EXAMINATION

The general physical examination in a delirious patient should include careful screening for signs of infection such as fever, tachypnea, pulmonary consolidation, heart murmur, and meningismus. The patient's fluid status should be assessed; both dehydration and fluid overload with resultant hypoxemia have been associated with delirium, and each is usually easily rectified. The appearance of the skin can be helpful, showing jaundice in hepatic encephalopathy, cyanosis in hypoxemia, or needle tracks in patients using intravenous drugs.

The neurologic examination requires a careful assessment of mental status. Patients with delirium often present with a fluctuating course; therefore, the diagnosis can be missed when one relies on a single time point of evaluation. For patients who worsen in the evening (sundowning), assessment only during morning rounds may be falsely reassuring.

An altered level of consciousness ranging from hyperarousal to lethargy to coma is present in most patients with delirium and can be assessed easily at the bedside. In a patient with a relatively normal level of consciousness, a screen for an attentional deficit is in order, because this deficit is the classic neuropsychological hallmark of delirium. Attention can be assessed while taking a history from the patient. Tangential speech, a fragmentary flow of ideas, or inability to follow complex commands often signifies an attentional problem. There are formal neuropsychological tests to assess attention, but a simple bedside test of digit span forward is quick and fairly sensitive. In this task, patients are asked to repeat successively longer random strings of digits beginning with two digits in a row, said to the patient at one per second intervals. Healthy adults can repeat a string of five to seven digits before faltering; a digit span of four or less usually indicates an attentional deficit unless hearing or language barriers are present, and many patients with delirium have digit spans of three or fewer digits.

More formal neuropsychological testing can be helpful in assessing a delirious patient, but it is usually too cumbersome and time-consuming in the inpatient setting. A Mini-Mental State Examination (MMSE) provides information regarding orientation, language, and visuospatial skills **(Chap. 29)**; however, performance of many tasks on the MMSE, including the spelling of "world" backward and serial subtraction of digits, will be impaired by delirious patients' attentional deficits, rendering the test unreliable.

The remainder of the screening neurologic examination should focus on identifying new focal neurologic deficits. Focal strokes or mass lesions in isolation are rarely the cause of delirium, but patients with underlying extensive cerebrovascular disease or neurodegenerative conditions may not be able to cognitively tolerate even relatively small new insults. Patients should be screened for other signs of neurodegenerative conditions such as parkinsonism, which is seen not only in idiopathic Parkinson's disease but also in other dementing conditions including Alzheimer's disease, dementia with Lewy bodies, and progressive supranuclear palsy. The presence of multifocal myoclonus or asterixis on the motor examination is nonspecific but usually indicates a metabolic or toxic etiology of the delirium.

ETIOLOGY

Some etiologies can be easily discerned through a careful history and physical examination, whereas others require confirmation with laboratory studies, imaging, or other ancillary tests. A large, diverse group of insults can lead to delirium, and the cause in many patients is multifactorial. Common etiologies are listed in **Table 27-2**.

Prescribed, over-the-counter, and herbal medications all can precipitate delirium. Drugs with anticholinergic properties, narcotics, and benzodiazepines are particularly common offenders, but nearly any compound can lead to cognitive dysfunction in a predisposed patient. Whereas an elderly patient with baseline dementia may become delirious upon exposure to a relatively low dose of a medication, in less susceptible individuals, delirium occurs only with very high doses of the same medication. This observation emphasizes the importance of correlating the timing of recent medication changes, including dose and formulation, with the onset of cognitive dysfunction.

In younger patients, illicit drugs and toxins are common causes of delirium. In addition to more classic drugs of abuse, the availability of "bath salts," synthetic cannabis **(Chap. 455)**, methylenedioxymethamphetamine (MDMA, ecstasy), γ-hydroxybutyrate (GHB), and the phencyclidine (PCP)-like agent ketamine has led to an increase in delirious young persons presenting to acute care settings **(Chap. 457)**. Many common prescription drugs such as oral narcotics and benzodiazepines are often abused and readily available on the street. Alcohol abuse leading to high serum levels causes confusion, but more commonly, it is withdrawal from alcohol that leads to a hyperactive delirium **(Chap. 453)**. Alcohol and benzodiazepine withdrawal should be considered in all cases of delirium, including in the elderly, because even patients who drink only a few servings of alcohol every day can experience relatively severe withdrawal symptoms upon hospitalization.

Metabolic abnormalities and underlying endocrinopathies such as thyroid and adrenal disorders.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Metabolic abnormalities such as electrolyte disturbances of sodium, calcium, magnesium, or glucose can cause delirium, and mild derangements can lead to substantial cognitive disturbances in susceptible individuals. Other common metabolic etiologies include liver and renal failure, hypercarbia and hypoxemia, vitamin deficiencies of thiamine and B_{12}, autoimmune disorders including central nervous system (CNS) vasculitis, and endocrinopathies such as thyroid and adrenal disorders.

Systemic infections often cause delirium, especially in the elderly. A common scenario involves the development of an acute cognitive decline in the setting of a urinary tract infection in a patient with baseline dementia. Pneumonia, skin infections such as cellulitis, and frank sepsis also lead to delirium. This so-called septic encephalopathy, often seen in the ICU, is probably due to the release of proinflammatory cytokines and their diffuse cerebral effects. CNS infections such as meningitis, encephalitis, and abscess are less common etiologies of delirium, as are cases of autoimmune or paraneoplastic encephalitis; however, in light of the high morbidity and mortality rates associated with these conditions when they are not treated, clinicians must always maintain a high index of suspicion.

In some susceptible individuals, exposure to the unfamiliar environment of a hospital itself can contribute to delirium. This etiology usually occurs as part of a multifactorial delirium and should be considered a diagnosis of exclusion after all other causes have been thoroughly investigated. Many primary prevention and treatment strategies for delirium involve relatively simple methods to address the aspects of the inpatient setting that are most confusing.

Cerebrovascular etiologies of delirium are usually due to global hypoperfusion in the setting of systemic hypotension from heart failure, septic shock, dehydration, or anemia. Focal strokes in the right parietal lobe and medial dorsal thalamus rarely can lead to a delirious state. A more common scenario involves a new focal stroke or hemorrhage causing confusion in a patient who has decreased cerebral reserve. In these individuals, it is sometimes difficult to distinguish between cognitive dysfunction resulting from the new neurovascular insult itself and delirium due to the infectious, metabolic, and pharmacologic complications that can accompany hospitalization after stroke.

Because a fluctuating course often is seen in delirium, intermittent seizures may be overlooked when one is considering potential etiologies. Both nonconvulsive status epilepticus and recurrent focal or generalized seizures followed by postictal confusion can cause delirium; EEG remains essential for this diagnosis and should be considered whenever the etiology of delirium remains unclear following initial workup. Seizure activity spreading from an electrical focus in a mass or infarct can explain global cognitive dysfunction caused by relatively small lesions.

It is extremely common for patients to experience delirium at the end of life in palliative care settings. This condition must be identified and treated aggressively because it is an important cause of patient and family discomfort at the end of life. It should be remembered that these patients also may be suffering from more common etiologies of delirium such as systemic infection.

LABORATORY AND DIAGNOSTIC EVALUATION

A cost-effective approach allows the history and physical examination to guide further tests. No single algorithm will fit all delirious patients due to the staggering number of potential etiologies, but one stepwise approach is detailed in **Table 27-3**. If a clear precipitant such as an offending medication is identified, further testing may not be required. If, however, no likely etiology is uncovered with initial evaluation, an aggressive search for an underlying cause should be initiated.

Basic screening labs, including a complete blood count, electrolyte panel, and tests of liver and renal function, should be obtained in all patients with delirium. In elderly patients, screening for systemic infection, including chest radiography, urinalysis and culture, and possibly blood cultures, is important. In younger individuals, serum and urine drug and toxicology screening may be appropriate earlier in the workup. Additional laboratory tests addressing other autoimmune, endocrinologic, metabolic, and infectious etiologies should be reserved for patients in whom the diagnosis remains unclear after initial testing.

Multiple studies have demonstrated that brain imaging in patients with delirium is often unhelpful. If, however, the initial workup is unrevealing, most clinicians quickly move toward imaging of the brain to exclude structural causes. A noncontrast computed tomography (CT) scan can identify large masses and hemorrhages but is otherwise unlikely to help determine an etiology of delirium. The ability of magnetic resonance imaging (MRI) to identify most acute ischemic strokes as well as to provide neuroanatomic detail that gives clues to possible infectious, inflammatory, neurodegenerative, and neoplastic conditions makes it the test of choice. Because MRI techniques are limited by availability, speed of imaging, patient's cooperation, and contraindications, many clinicians begin with CT scanning and proceed to MRI if the etiology of delirium remains elusive.

Lumbar puncture (LP) must be obtained immediately after neuroimaging for all patients in whom CNS infection is suspected. Spinal fluid

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 27-1

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Lumbar puncture (LP) must be obtained immediately after neuroimaging for all patients in whom CNS infection is suspected. Spinal fluid examination can also be useful in identifying autoimmune, other inflammatory, and neoplastic conditions. As a result, LP should be considered in any delirious patient with a negative workup. EEG remains invaluable if seizures are considered or if there is no cause readily identified.

TABLE 27-1

The Confusion Assessment Method (CAM) Diagnostic Algorithm[a]

The diagnosis of delirium requires the presence of features 1 and 2 **and** *either* feature 3 or 4.
Feature 1. Acute Onset and Fluctuating Course
This feature is satisfied by positive responses to the following questions: Is there evidence of an acute change in mental status from the patient's baseline? Did the (abnormal) behavior fluctuate during the day, that is, tend to come and go, or did it increase and decrease in severity?
Feature 2. Inattention
This feature is satisfied by a positive response to the following question: Did the patient have difficulty focusing attention, for example, being easily distractible, or have difficulty keeping track of what was being said?
Feature 3. Disorganized Thinking
This feature is satisfied by a positive response to the following question: Was the patient's thinking disorganized or incoherent, such as rambling or irrelevant conversation, unclear or illogical flow of ideas, or unpredictable switching from subject to subject?
Feature 4. Altered Level of Consciousness
This feature is satisfied by any answer other than "alert" to the following question: Overall, how would you rate the patient's level of consciousness: alert (normal), vigilant (hyperalert), lethargic (drowsy, easily aroused), stupor (difficult to arouse), or coma (unarousable)?

[a]Information is usually obtained from a reliable reporter, such as a family member, caregiver, or nurse.

Source: From Annals of Internal Medicine. SK Inouye et al: Clarifying confusion: The Confusion Assessment Method. A new method for detection of delirium. 113:941, 1990. Copyright © 1990 American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.

TABLE 27-2

Differential Diagnosis of Delirium

Toxins
Prescription medications: especially those with anticholinergic properties, narcotics, and benzodiazepines
Drugs of abuse: alcohol intoxication and alcohol withdrawal, opiates, ecstasy, LSD, GHB, PCP, ketamine, cocaine, "bath salts," marijuana and its synthetic forms
Poisons: inhalants, carbon monoxide, ethylene glycol, pesticides
Metabolic Conditions
Electrolyte disturbances: hypoglycemia, hyperglycemia, hyponatremia, hypernatremia, hypercalcemia, hypocalcemia, hypomagnesemia
Liver failure/hepatic encephalopathy

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR

Electrolyte disturbances: hypoglycemia, hyperglycemia, hyponatremia, hypernatremia, hypercalcemia, hypocalcemia, hypomagnesemia
Hypothermia and hyperthermia
Pulmonary failure: hypoxemia and hypercarbia
Liver failure/hepatic encephalopathy
Renal failure/uremia
Cardiac failure
Vitamin deficiencies: B_{12}, thiamine, folate, niacin
Dehydration and malnutrition
Anemia
Infections
Systemic infections: urinary tract infections, pneumonia, skin and soft tissue infections, sepsis
CNS infections: meningitis, encephalitis, brain abscess
Endocrine Conditions
Hyperthyroidism, hypothyroidism
Hyperparathyroidism
Adrenal insufficiency
Cerebrovascular Disorders
Global hypoperfusion states
Hypertensive encephalopathy
Focal ischemic strokes and hemorrhages (rare): especially nondominant parietal and thalamic lesions
Autoimmune Disorders
CNS vasculitis
Cerebral lupus
Neurologic paraneoplastic and autoimmune encephalitis
Seizure-Related Disorders
Nonconvulsive status epilepticus
Intermittent seizures with prolonged postictal states

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Gliomatosis cerebri

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Neoplastic Disorders
Diffuse metastases to the brain
Gliomatosis cerebri
Carcinomatous meningitis
CNS lymphoma
Hospitalization
Terminal end-of-life delirium

Abbreviations: CNS, central nervous system; GHB, γ-hydroxybutyrate; LSD, lysergic acid diethylamide; PCP, phencyclidine.

TABLE 27-3

Stepwise Evaluation of a Patient with Delirium

Initial Evaluation
History with special attention to medications (including over-the-counter and herbals)
General physical examination and neurologic examination
Complete blood count
Electrolyte panel including calcium, magnesium, phosphorus
Liver function tests, including albumin
Renal function tests
First-Tier Further Evaluation Guided by Initial Evaluation
Systemic infection screen
Urinalysis and culture
Chest radiograph
Blood cultures
Electrocardiogram
Arterial blood gas
Serum and/or urine toxicology screen (perform earlier in young persons)
Brain imaging with MRI with diffusion and gadolinium (preferred) or CT
Second-Tier Further Evaluation

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Suspected CNS infection or other inflammatory disorder: lumbar puncture after brain imaging
Suspected seizure-related etiology: electroencephalogram (EEG) (if high suspicion, should be performed immediately)
Second-Tier Further Evaluation
Vitamin levels: B_{12}, folate, thiamine
Endocrinologic laboratories: thyroid-stimulating hormone (TSH) and free T_4; cortisol
Serum ammonia
Sedimentation rate
Autoimmune serologies: antinuclear antibodies (ANA), complement levels; p-ANCA, c-ANCA, consider paraneoplastic/autoimmune encephalitis serologies
Infectious serologies: rapid plasmin reagin (RPR); fungal and viral serologies if high suspicion; HIV antibody
Lumbar puncture (if not already performed)
Brain MRI with and without gadolinium (if not already performed)

Abbreviations: c-ANCA, cytoplasmic antineutrophil cytoplasmic antibody; CNS, central nervous system; CT, computed tomography; MRI, magnetic resonance imaging; p-ANCA, perinuclear antineutrophil cytoplasmic antibody.

TREATMENT

Delirium

Management of delirium begins with treatment of the underlying inciting factor (e.g., patients with systemic infections should be given appropriate antibiotics, and underlying electrolyte disturbances should be judiciously corrected). These treatments often lead to prompt resolution of delirium. Blindly targeting the symptoms of delirium pharmacologically only serves to prolong the time patients remain in the confused state and may mask important diagnostic information.

Relatively simple methods of supportive care can be highly effective **(Fig. 27-1)**. Reorientation by the nursing staff and family combined with visible clocks, calendars, and outside-facing windows can reduce confusion. Sensory isolation should be prevented by providing glasses and hearing aids to patients who need them. Sundowning can be addressed to a large extent through vigilance to appropriate sleep-wake cycles. During the day, a well-lit room should be accompanied by activities or exercises to prevent napping. At night, a quiet, dark environment with limited interruptions by staff can assure proper rest; melatonin can be considered before bed to promote sleep. These sleep-wake cycle interventions are especially important in the ICU setting as the usual constant 24-h activity commonly provokes delirium. Attempting to mimic the home environment as much as possible also has been shown to help treat and even prevent delirium. Visits from friends and family throughout the day minimize the anxiety associated with the constant flow of new faces of staff and physicians. Allowing hospitalized patients to have access to home bedding, clothing, and nightstand objects makes the hospital environment less foreign and therefore less confusing. Simple standard nursing practices such as maintaining proper nutrition and volume status as well as managing pain, incontinence, and skin breakdown also help alleviate discomfort and resulting confusion.

In some instances, patients pose a threat to their own safety or to the safety of staff members, and acute management is required. Bed alarms and personal sitters are more effective and much less disorienting than physical restraints. Chemical restraints should be avoided, but when necessary, very-low-dose typical or atypical antipsychotic medications administered on an as-needed basis can be used, recognizing that clinical trials have consistently shown that these medications are ineffective in treating delirium. Therefore, they should be reserved for patients who display severe

caused by alcohol or benzodiazepine withdrawal.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Delirium

Management of delirium begins with treatment of the underlying inciting factor (e.g., patients with systemic infections should be given appropriate antibiotics, and underlying electrolyte disturbances should be judiciously corrected). These treatments often lead to prompt resolution of delirium. Blindly targeting the symptoms of delirium pharmacologically only serves to prolong the time patients remain in the confused state and may mask important diagnostic information.

Relatively simple methods of supportive care can be highly effective **(Fig. 27-1)**. Reorientation by the nursing staff and family combined with visible clocks, calendars, and outside-facing windows can reduce confusion. Sensory isolation should be prevented by providing glasses and hearing aids to patients who need them. Sundowning can be addressed to a large extent through vigilance to appropriate sleep-wake cycles. During the day, a well-lit room should be accompanied by activities or exercises to prevent napping. At night, a quiet, dark environment with limited interruptions by staff can assure proper rest; melatonin can be considered before bed to promote sleep. These sleep-wake cycle interventions are especially important in the ICU setting as the usual constant 24-h activity commonly provokes delirium. Attempting to mimic the home environment as much as possible also has been shown to help treat and even prevent delirium. Visits from friends and family throughout the day minimize the anxiety associated with the constant flow of new faces of staff and physicians. Allowing hospitalized patients to have access to home bedding, clothing, and nightstand objects makes the hospital environment less foreign and therefore less confusing. Simple standard nursing practices such as maintaining proper nutrition and volume status as well as managing pain, incontinence, and skin breakdown also help alleviate discomfort and resulting confusion.

In some instances, patients pose a threat to their own safety or to the safety of staff members, and acute management is required. Bed alarms and personal sitters are more effective and much less disorienting than physical restraints. Chemical restraints should be avoided, but when necessary, very-low-dose typical or atypical antipsychotic medications administered on an as-needed basis can be used, recognizing that clinical trials have consistently shown that these medications are ineffective in treating delirium. Therefore, they should be reserved for patients who display severe agitation and significant potential to harm themselves or staff. The association of antipsychotic use in the elderly with increased mortality rates underscores the importance of using these medications judiciously and only as a last resort. Benzodiazepines often worsen confusion through their sedative properties. Although many clinicians use benzodiazepines to treat acute confusion, their use should be limited to cases in which delirium is caused by alcohol or benzodiazepine withdrawal.

FIGURE 27-1

Delirium management and prevention: a checklist for hospitalized patients. Effective management of delirium relies on broad efforts to promote wakefulness (***A***) and sleep (***B***). CPO, continuous pulse oximetry.



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

promote wakefulness (**A**) and sleep (**B**). CPO, continuous pulse oximetry.



PREVENTION

In light of the high morbidity associated with delirium and the tremendously increased health care costs that accompany it, development of an effective strategy to prevent delirium in hospitalized patients is extremely important. Successful identification of high-risk patients is the first step, followed by initiation of appropriate interventions. Increasingly, hospitals are using nursing or physician-administered tools to screen for high-risk individuals, triggering simple standardized protocols used to manage risk factors for delirium, including sleep-wake cycle reversal, immobility, visual impairment, hearing impairment, sleep deprivation, and dehydration. No specific medications have been definitively shown to be effective for delirium prevention, including trials of cholinesterase inhibitors and antipsychotic agents. Melatonin and its agonist ramelteon have shown some promising results in small preliminary trials. Recent studies in the ICU have focused both on identifying sedatives, such as dexmedetomidine, that are less likely to lead to

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FURTHER READING

Access Provided by:
SILVERCHAIR

preliminary trials. Recent studies in the ICU have focused both on identifying sedatives, such as dexmedetomidine, that are less likely to lead to delirium in critically ill patients and on developing protocols for daily awakenings in which infusions of sedative medications are interrupted and the patient is reorientated by the staff. All hospitals and health care systems should work toward decreasing the incidence of delirium and promptly recognizing and treating the disorder when it occurs.

FURTHER READING

Brown EG et al: Evaluation of a multicomponent pathway to address inpatient delirium on a neurosciences ward. BMC Health Serv Res 18:106, 2018. [PubMed: 29433572]

Constantin JM et al: Efficacy and safety of sedation with dexmedetomidine in critical care patients: A meta-analysis of randomized controlled trials. Anaesth Crit Care Pain Med 35:7, 2016. [PubMed: 26700947]

Girard TD et al: Haloperidol and ziprasidone for treatment of delirium in critical illness. N Engl J Med 379:2506, 2018. [PubMed: 30346242]

Goldberg TE et al: Association of delirium with long-term cognitive decline: A meta-analysis. JAMA Neurol 77:1, 2020.

Hatta K et al: Preventive effects of ramelteon on delirium: A randomized placebo-controlled trial. JAMA Psychiatry 71:397, 2014. [PubMed: 24554232]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 27: Confusion and Delirium, S. Andrew Josephson; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 28: Coma

S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser

INTRODUCTION

Coma is among the most common neurologic emergencies encountered general medicine and requires an organized approach. It accounts for a substantial portion of admissions to emergency wards and occurs on all hospital services.

There exists a continuum of states of reduced alertness, the most severe form being coma, defined as a deep sleeplike state with eyes closed, from which the patient cannot be aroused. Stupor refers to a lower threshold for arousability, in which the patient can be transiently awakened by vigorous stimuli, accompanied by motor behavior that leads to avoidance or withdrawal from noxious stimuli. Drowsiness simulates light sleep and is characterized by easy arousal that may persist for brief periods. Stupor and drowsiness are usually accompanied by some degree of confusion when the patient is alerted **(Chap. 27)**. A precise narrative description of the level of arousal and of the type of responses evoked by various stimuli as observed at the bedside is preferable to use of ambiguous terms such as lethargy, semicoma, or obtundation.

Several conditions that render patients unresponsive and simulate coma are considered separately because of their special significance. The *vegetative state* signifies an awake-appearing but nonresponsive state, usually encountered in a patient who has emerged from coma. In the vegetative state, the eyelids may open periodically, giving the appearance of wakefulness. Respiratory and autonomic functions are retained. Yawning, coughing, swallowing, and limb and head movements persist, but there are few, if any, meaningful responses to the external and internal environment. There are typically accompanying signs that indicate extensive damage in both cerebral hemispheres, e.g., decerebrate or decorticate limb posturing and absent responses to visual stimuli (see below). In the closely related but less severe *minimally conscious state*, the patient displays rudimentary vocal or motor behaviors, often spontaneous, but sometimes in response to touch, visual stimuli, or command. Cardiac arrest with cerebral hypoperfusion and head trauma are the most common causes of the vegetative and minimally conscious states **(Chap. 307)**.

The prognosis for regaining meaningful mental faculties once the vegetative state has supervened for several months is poor, and after a year, almost nil; hence the term *persistent vegetative state*. Most reports of dramatic recovery, when investigated carefully, are found to yield to the usual rules for prognosis, but there have been rare instances in which recovery has occurred to a severely disabled condition and, in rare childhood cases, to an even better state. Patients in the minimally conscious state carry a better prognosis for some recovery compared to those in a persistent vegetative state, but even in these patients, dramatic recovery after 12 months is unusual.

The possibility of incorrectly attributing meaningful behavior to patients in the vegetative and minimally conscious states creates problems and anguish for families and physicians. The question of whether some of these patients have the capability for cognition has been investigated by functional MRI and electroencephalogram (EEG) studies that have demonstrated cerebral activation that is temporally consistent in response to verbal and other stimuli, as discussed in more detail below. This finding suggests at a minimum that some of these patients could in the future be able to communicate their needs using technological advances and that further research could shed light on treatment approaches targeting areas of the brain and their connections that seem to be preserved in individual patients.

Several syndromes that affect alertness are prone to be misinterpreted as stupor or coma, and clinicians should be aware of these pitfalls when diagnosing coma at the bedside. Akinetic mutism refers to a partially or fully awake state in which the patient remains virtually immobile and mute but can form impressions and think, as demonstrated by later recounting of events. This condition results from damage in the regions of the medial thalamic nuclei or the frontal lobes (particularly lesions situated deeply or on the orbitofrontal surfaces) or from extreme hydrocephalus. The term *abulia* describes a milder form of akinetic mutism characterized by mental and physical slowness and diminished ability to initiate activity. It is also usually the result of damage to the medial frontal lobes and their connections **(Chap. 30)**.

Catatonia is a hypomobile and mute syndrome that occurs usually as part of a major psychosis, typically schizophrenia or major depression. Catatonic patients make few voluntary or responsive movements, although they blink, swallow, and may not appear distressed. There are nevertheless signs that

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

With recovery from catatonia, patients often have some memory of events that occurred during their stupor. Catatonia is superficially similar to

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Catatonia is a hypomobile and mute syndrome that occurs usually as part of a major psychosis, typically schizophrenia or major depression. Catatonic patients make few voluntary or responsive movements, although they blink, swallow, and may not appear distressed. There are nevertheless signs that the patient is responsive, although it takes a careful examination to demonstrate these features. For example, eyelid elevation is actively resisted, blinking occurs in response to a visual threat, and the eyes move concomitantly with head rotation, all of which are inconsistent with the presence of a brain lesion causing unresponsiveness. The limbs may retain postures in which they have been placed by the examiner ("waxy flexibility," or catalepsy). With recovery from catatonia, patients often have some memory of events that occurred during their stupor. Catatonia is superficially similar to akinetic mutism, but clinical evidence of cerebral damage such as hyperreflexia and hypertonicity of the limbs is lacking in the former. The special problem of coma in brain death is discussed below.

The locked-in state describes a type of pseudocoma in which an awake but paralyzed patient has no means of producing speech or volitional limb movement but retains voluntary vertical eye movements and lid elevation, thus allowing the patient to communicate. The pupils are normally reactive. The usual cause is an infarction (e.g., basilar artery thrombosis) or hemorrhage of the bilateral ventral pons that transects all descending motor (corticospinal and corticobulbar) pathways. Another awake but de-efferented state occurs as a result of total paralysis of the musculature in severe cases of neuromuscular weakness such as in Guillain-Barré syndrome **(Chap. 447)**, critical illness neuropathy **(Chap. 307)**, or pharmacologic neuromuscular blockade.

THE ANATOMY AND PHYSIOLOGY OF COMA

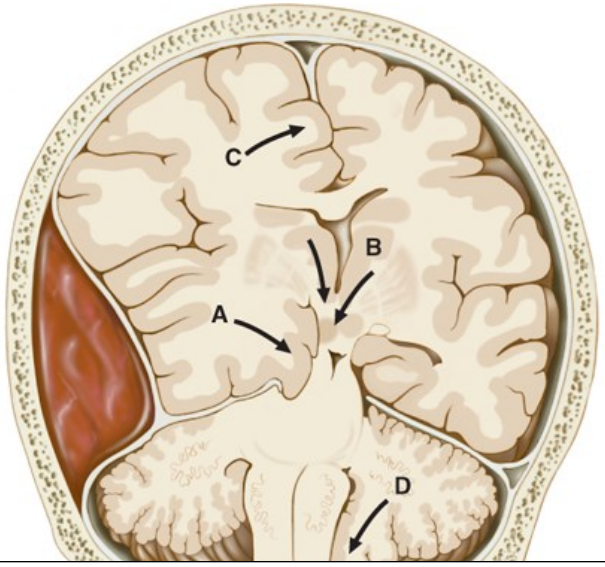
Almost all instances of coma can be traced to either (1) widespread abnormalities of the cerebral hemispheres or (2) reduced activity of the thalamocortical alerting system, the reticular activating system (RAS), which is an assemblage of neurons located diffusely in the upper brainstem and thalamus. The proper functioning of this system, its ascending projections to the cortex, and the cortex itself are required to maintain alertness and coherence of thought. In addition to structural damage to either or both of these systems, suppression of reticulocerebral function commonly occurs by drugs, toxins, or metabolic derangements such as hypoglycemia, anoxia, uremia, and hepatic failure, or by seizures; these types of metabolic causes of coma are far more common than structural injuries.

Coma Due to Cerebral Mass Lesions and Herniation Syndromes

The skull prevents outward expansion of the brain, and infoldings of the dura create compartments that restrict displacement of brain tissue within the cranium. The two cerebral hemispheres are separated by the falx and the anterior and posterior fossae by the tentorium. Herniation refers to displacement of brain tissue by an intracerebral or overlying mass into a contiguous compartment that it normally does not occupy. Coma from mass lesions, and many of its associated signs, are attributable to these tissue shifts, and certain clinical features are characteristic of specific configurations of herniation **(Fig. 28-1)**.

FIGURE 28-1

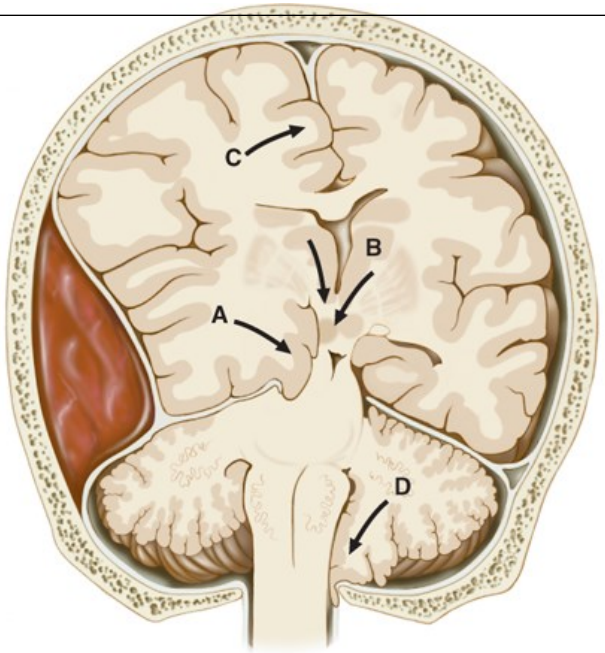
Types of cerebral herniation: (***A***) uncal; (***B***) central; (***C***) transfalcial; and (***D***) foraminal.



Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

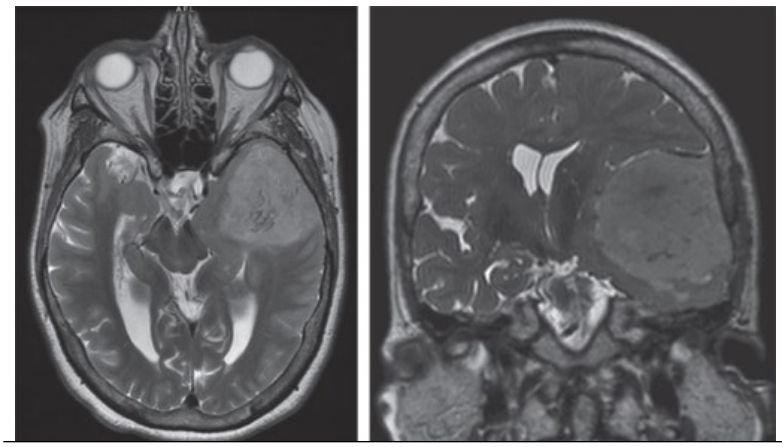


Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

In the most common form of herniation, brain tissue is displaced from the supratentorial to the infratentorial compartment through the tentorial opening, referred to as transtentorial herniation. The cause is often a mass hemispheral lesion, with accompanying contralateral hemiparesis. Uncal transtentorial herniation refers to impaction of the anterior medial temporal gyrus (the uncus) into the tentorial opening just anterior to and adjacent to the midbrain **(Fig. 28-1A)**. The uncus can compress the third nerve as the nerve traverses the subarachnoid space, causing enlargement of the ipsilateral pupil as the first sign (the fibers subserving parasympathetic pupillary function are located peripherally in the nerve). The coma that typically follows is due to lateral displacement of the midbrain (and therefore the RAS) against the opposite tentorial edge by the displaced parahippocampal gyrus **(Fig. 28-2)**, compressing the opposite cerebral peduncle and producing a Babinski sign and ipsilateral hemiparesis (the Kernohan-Woltman sign). Herniation may also compress the anterior and posterior cerebral arteries as they pass over the tentorial reflections, with resultant brain infarction. These distortions may also entrap portions of the ventricular system, causing hydrocephalus.

FIGURE 28-2

Axial (A) and coronal (B) T2-weighted magnetic resonance images from a stuporous patient with a left third nerve palsy from a large left-sided meningioma. A. The upper midbrain is compressed and displaced horizontally away from the mass, and there is transtentorial herniation of the medial temporal lobe structures, including the uncus. **B.** The lateral ventricle opposite to the mass has become enlarged as a result of compression of the third ventricle. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



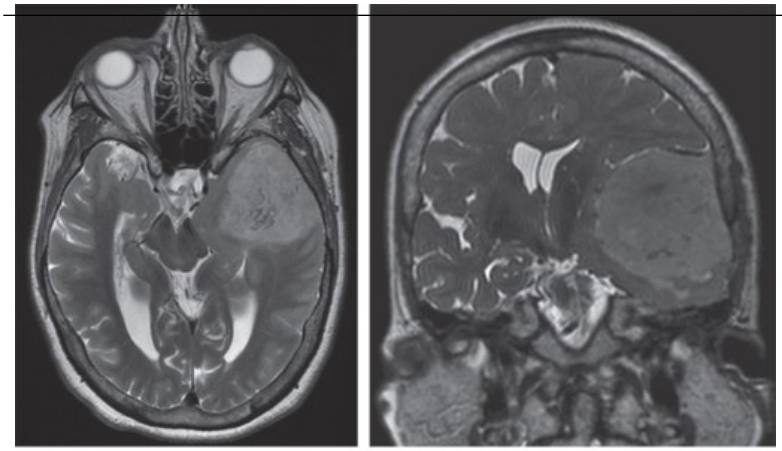
Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Central transtentorial herniation denotes a symmetric downward movement of the thalamic structures through the tentorial opening with compression of the upper midbrain **(Fig. 28-1B)**. Miotic pupils and drowsiness are the heralding signs, in contrast to a unilaterally enlarged pupil of

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



A B

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Central transtentorial herniation denotes a symmetric downward movement of the thalamic structures through the tentorial opening with compression of the upper midbrain **(Fig. 28-1B)**. Miotic pupils and drowsiness are the heralding signs, in contrast to a unilaterally enlarged pupil of the uncal syndrome. Both uncal and central transtentorial herniations cause progressive compression of the brainstem and RAS, with initial damage to the midbrain, then the pons, and finally the medulla. The result is an approximate sequence of neurologic signs that corresponds to each affected level, with respiratory centers in the brainstem often spared until late in the herniation syndrome. Other forms of herniation include transfalcial herniation (displacement of the cingulate gyrus under the falx and across the midline, **Fig. 28-1C**) and foraminal herniation (downward forcing of the cerebellar tonsils into the foramen magnum, **Fig. 28-1D**), which causes early compression of the medulla, respiratory arrest, and death.

Coma Due to Metabolic, Drug, and Toxic Disorders

Many systemic metabolic abnormalities cause coma by interrupting the delivery of energy substrates (e.g., oxygen, glucose) or by altering neuronal excitability (drugs and alcohol, anesthesia, and epilepsy). These are the most common causes of coma in large case series. The metabolic abnormalities that produce coma may, in milder forms, induce a confusional state (metabolic encephalopathy) in which clouded consciousness and coma are in a continuum.

Cerebral neurons are dependent on cerebral blood flow (CBF) and the delivery of oxygen and glucose. Brain stores of glucose are able to provide energy for ~2 min after blood flow is interrupted, and oxygen stores last 8–10 s after the cessation of blood flow. Simultaneous hypoxia and ischemia exhaust glucose more rapidly. The EEG rhythm in these circumstances becomes diffusely slowed, typical of metabolic encephalopathies, and as substrate delivery worsens, eventually brain electrical activity ceases.

Unlike hypoxia-ischemia, which first causes a metabolic encephalopathy due to reduced energy substrate but ultimately causes neuronal destruction, most metabolic disorders such as hypoglycemia, hyponatremia, hyperosmolarity, hypercapnia, hypercalcemia, and hepatic and renal failure cause no or only minor neuropathologic changes in the brain. The reversible effects of these conditions are not fully understood but may result from impaired energy supplies, changes in ion fluxes across neuronal membranes, and neurotransmitter abnormalities. In hepatic encephalopathy (HE), high ammonia concentrations lead to increased synthesis of glutamine in astrocytes and osmotic swelling of the cells, mitochondrial energy failure, production of reactive nitrogen and oxygen species, increases in the inhibitory neurotransmitter GABA, and synthesis of putative "false" neurotransmitters. Over time, development of a diffuse astrocytosis is typical of chronic HE. Which, if any, of these is responsible for coma is not known.

The mechanism of the encephalopathy of renal failure is also uncertain and likely to be multifactorial; unlike ammonia, urea does not produce central nervous system (CNS) depression. Contributors to uremic encephalopathy may include accumulation of neurotoxic substances such as creatinine, guanidine, and related compounds; depletion of catecholamines; altered glutamate and GABA tone; increases in brain calcium; inflammation with disruption of the blood-brain barrier; and frequent coexisting vascular disease.

Coma and seizures are common accompaniments of large shifts in sodium and water balance in the brain. These changes in osmolarity arise from systemic medical disorders, including diabetic ketoacidosis, the nonketotic hyperosmolar state, and hyponatremia from any cause (e.g., water

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

these metabolic encephalopathies, the degree of neurologic change depends on the rapidity with which the serum changes occur. The pathophysiology of other metabolic encephalopathies such as those due to hypercalcemia, hypothyroidism, vitamin B_{12} deficiency, and hypothermia

systemic medical disorders, including diabetic ketoacidosis, the nonketotic hyperosmolar state, and hyponatremia from any cause (e.g., water intoxication, excessive secretion of antidiuretic hormone, or atrial natriuretic peptides). Sodium levels <125 mmol/L, especially if achieved quickly, induce confusion, and levels <119 mmol/L are typically associated with coma and convulsions. In hyperosmolar coma, the serum osmolarity is generally >350 mosmol/L. Hypercapnia depresses the level of consciousness in proportion to the rise in carbon dioxide (CO_2) in the blood. In all of these metabolic encephalopathies, the degree of neurologic change depends on the rapidity with which the serum changes occur. The pathophysiology of other metabolic encephalopathies such as those due to hypercalcemia, hypothyroidism, vitamin B_{12} deficiency, and hypothermia are incompletely understood but must reflect derangements of CNS biochemistry, membrane function, or neurotransmitters.

Comas due to drugs and toxins are typically reversible and leave no residual damage provided there has not been hypoxia or severe hypotension. Many drugs and toxins are capable of depressing nervous system function. Some produce coma by affecting both the RAS and the cerebral cortex. The combination of cortical and brainstem signs, which occurs occasionally in certain drug overdoses, may lead to an incorrect diagnosis of structural brainstem disease. Overdose of medications that have atropinic actions produces signs such as dilated pupils, tachycardia, and dry skin; opiate overdose produces pinpoint pupils <1 mm in diameter. Some drug intoxications, typified by barbiturates, can mimic all of the signs of brain death; thus, toxic etiologies should be excluded prior to making a diagnosis of brain death.

Epileptic Coma

Generalized electrical seizures are associated with coma, even in the absence of motor convulsions (nonconvulsive status epilepticus). As a result, EEG monitoring is often used in the evaluation of unexplained coma to exclude this treatable etiology. The self-limited coma that follows a seizure, the postictal state, may be due to exhaustion of energy reserves or effects of locally toxic molecules that are the by-product of seizures. The postictal state produces continuous, generalized slowing of the background EEG activity similar to that of metabolic encephalopathies. It typically lasts for a few minutes but in some cases can be prolonged for hours or even rarely for days.

Coma Due to Widespread Structural Damage to the Cerebral Hemispheres

This category, comprising several unrelated disorders, results from extensive bilateral structural cerebral damage. The clinical appearance simulates a metabolic encephalopathy. Hypoxia-ischemia is perhaps the best characterized form of this type of injury, in which it is not possible initially to distinguish the acute reversible effects of oxygen deprivation of the brain from the subsequent effects of anoxic neuronal damage. Similar cerebral damage may be produced by disorders that occlude widespread small blood vessels throughout the brain; examples include thrombotic thrombocytopenic purpura, hyperviscosity, and cerebral malaria. Diffuse white matter damage from cranial trauma or inflammatory demyelinating diseases can cause a similar coma syndrome.

APPROACH TO THE PATIENT

Coma

A video examination of the comatose patient is shown in Chap. V4. Acute respiratory and cardiovascular problems should be attended to prior to neurologic assessment. In most instances, a complete medical evaluation, except for vital signs, funduscopy, and examination for nuchal rigidity, may be deferred until the neurologic evaluation has established the severity and nature of coma. **The approach to the patient with coma from cranial trauma is discussed in Chap. 443.**

HISTORY

The cause of coma may be immediately evident as in cases of trauma, cardiac arrest, or observed drug ingestion. In the remainder, certain points are useful: (1) the circumstances and rapidity with which neurologic symptoms developed; (2) antecedent symptoms (confusion, weakness, headache, fever, seizures, dizziness, double vision, or vomiting); (3) the use of medications, drugs, or alcohol; and (4) chronic liver, kidney, lung, heart, or other medical disease. Direct interrogation of family, observers, and emergency medical technicians on the scene, in person or by telephone, is an important part of the evaluation when possible.

GENERAL PHYSICAL EXAMINATION

Signs of head trauma raise the possibility of coexisting spinal cord injury, and in such cases, immobilization of the cervical spine is essential to prevent further injury. Fever suggests a systemic infection, bacterial meningitis, encephalitis, heat stroke, neuroleptic malignant syndrome, malignant hyperthermia due to anesthetics, or anticholinergic drug intoxication. Only rarely is fever attributable to a lesion that has disturbed itself causes coma when the temperature is <31°C (87.8°F) regardless of the underlying etiology; less dramatically low body temperatures can also cause coma in some instances. Tachypnea may indicate systemic acidosis or pneumonia. Aberrant respiratory patterns that reflect brainstem

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

malignant hyperthermia due to anesthetics, or anticholinergic drug intoxication. Only rarely is fever attributable to a lesion that has disturbed hypothalamic temperature-regulating centers ("central fever"), and this diagnosis should only be considered after an exhaustive search for other causes fails to reveal an explanation for fever. A slight elevation in temperature may follow vigorous convulsions. Hypothermia is observed with alcohol, barbiturate, sedative, or phenothiazine intoxication; hypoglycemia; peripheral circulatory failure; or extreme hypothyroidism. Hypothermia itself causes coma when the temperature is <31°C (87.8°F) regardless of the underlying etiology; less dramatically low body temperatures can also cause coma in some instances. Tachypnea may indicate systemic acidosis or pneumonia. Aberrant respiratory patterns that reflect brainstem disorders are discussed below. Marked hypertension suggests hypertensive encephalopathy, cerebral hemorrhage, large cerebral infarction, or head injury. Hypotension is characteristic of coma from alcohol or barbiturate intoxication, internal hemorrhage or myocardial infarction causing poor delivery of blood to the brain, sepsis, profound hypothyroidism, or Addisonian crisis. The funduscopic examination can detect increased intracranial pressure (ICP) (papilledema), subarachnoid hemorrhage (subhyaloid hemorrhages), and hypertensive encephalopathy (exudates, hemorrhages, vessel-crossing changes, papilledema). Cutaneous petechiae suggest thrombotic thrombocytopenic purpura, meningococcemia, or a bleeding diathesis associated with an intracerebral hemorrhage. Cyanosis and reddish or anemic skin coloration are other indications of an underlying systemic disease or carbon monoxide as responsible for the coma.

NEUROLOGIC EXAMINATION

The patient should first be observed without intervention by the examiner. Spontaneously moving about the bed, reaching up toward the face, crossing legs, yawning, swallowing, coughing, and moaning reflect a drowsy state that is close to normal awakeness. Lack of restless movements on one side or an outturned leg suggests hemiplegia. Subtle, intermittent twitching movements of a foot, finger, or facial muscle may be the only sign of seizures. Multifocal myoclonus usually indicates a metabolic disorder, particularly uremia, anoxia, drug intoxication, or rarely a prion disease **(Chap. 438)**. In a drowsy and confused patient, bilateral asterixis is a sign of metabolic encephalopathy or drug intoxication.

Decorticate rigidity and decerebrate rigidity, or "posturing," describe stereotyped arm and leg movements occurring spontaneously or elicited by sensory stimulation. Flexion of the elbows and wrists and supination of the arm (decorticate posturing) classically suggest bilateral damage rostral to the midbrain, whereas extension of the elbows and wrists with pronation (decerebrate posturing) indicates damage to motor tracts caudal to the midbrain. However, these localizations have been adapted from animal work and cannot be applied with precision to coma in humans. In fact, acute and widespread disorders of any type, regardless of location, frequently cause limb extension.

LEVEL OF AROUSAL

A sequence of increasingly intense stimuli is first used to determine the threshold for arousal and the motor response of each side of the body. The results of testing may vary from minute to minute, and serial examinations are useful. Tickling the nostrils with a cotton wisp is a moderate stimulus to arousal—all but deeply stuporous and comatose patients will move the head away and arouse to some degree. An even greater degree of responsiveness is present if the patient uses his hand to remove an offending stimulus. Pressure on bony prominences and pinprick stimulation, when necessary, are humane forms of noxious stimuli; pinching the skin causes ecchymoses and is generally not performed but may be useful in eliciting abduction withdrawal movements of the limbs. Posturing in response to noxious stimuli indicates severe damage to the corticospinal system, whereas abduction-avoidance movement of a limb is usually purposeful and denotes an intact corticospinal system. Posturing may also be unilateral and coexist with purposeful limb movements, reflecting incomplete damage to the motor system.

BRAINSTEM REFLEXES

Assessment of brainstem function is essential to localization of the lesion in coma **(Fig. 28-3)**. Patients with preserved brainstem reflexes typically have a bihemispheric localization to coma, including toxic or drug intoxication, whereas patients with abnormal brainstem reflexes either have a lesion in the brainstem or a herniation syndrome from a cerebral mass lesion impacting the brainstem secondarily. The most important brainstem reflexes are pupillary size and reaction to light, spontaneous and elicited eye movements, corneal responses, and the respiratory pattern.

Pupillary Signs

Pupillary reactions are examined with a bright, diffuse light. Reactive and round pupils of midsize (2.5–5 mm) essentially exclude upper midbrain damage, either primary or secondary to compression from herniation. A response to light may be difficult to appreciate in pupils <2 mm in diameter, and bright room lighting may mute pupillary reactivity. One enlarged (>6 mm) and poorly reactive pupil signifies compression of the third nerve from the effects of a cerebral mass above. Enlargement of the pupil contralateral to a hemispheral mass may occur but is infrequent. An oval and slightly eccentric pupil is a transitional sign that accompanies early midbrain–third nerve compression. The most extreme pupillary sign, bilaterally dilated and unreactive pupils, indicates severe midbrain damage, usually from compression by a supratentorial mass. Ingestion of drugs with anticholinergic activity, the use of mydriatic eye drops, nebulizer treatments, and direct ocular trauma are other causes of pupillary enlargement.

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Unilateral miosis in coma has been attributed to dysfunction of sympathetic efferents originating in the posterior hypothalamus and descending in the tegmentum of the brainstem to the cervical cord. It is an occasional finding in patients with a large cerebral hemorrhage that affects the

Access Provided by:
SILVERCHAIR

Reactive and bilaterally small (1–2.5 mm) but not pinpoint pupils are seen in metabolic encephalopathies or in deep bilateral hemispheral lesions such as hydrocephalus or thalamic hemorrhage. Even smaller reactive pupils (<1 mm) characterize opioid overdoses but also occur with extensive pontine hemorrhage. The response to naloxone and the presence of reflex eye movements (see below) assist in distinguishing between these. Unilateral miosis in coma has been attributed to dysfunction of sympathetic efferents originating in the posterior hypothalamus and descending in the tegmentum of the brainstem to the cervical cord. It is an occasional finding in patients with a large cerebral hemorrhage that affects the thalamus.

Ocular Movements

The eyes are first observed by elevating the lids and observing the resting position and spontaneous movements of the globes. Horizontal divergence of the eyes at rest is normal in drowsiness. As coma deepens, the ocular axes may become parallel again.

Spontaneous eye movements in coma often take the form of conjugate horizontal roving. This finding alone exonerates extensive damage in the midbrain and pons and has the same significance as normal reflex eye movements (see below). Conjugate horizontal ocular deviation to one side indicates damage to the frontal lobe on the same side or less commonly the pons on the opposite side. This phenomenon is summarized by the following maxim: *The eyes look toward a hemispheral lesion and away from a brainstem lesion.* Seizures involving the frontal lobe drive the eyes to the opposite side, simulating a pontine destructive lesion. The eyes may occasionally turn paradoxically away from the side of a deep hemispheral lesion (“wrong-way eyes”). The eyes turn down and inward with thalamic and upper midbrain lesions, typically thalamic hemorrhage. “Ocular bobbing” describes brisk downward and slow upward movements of the eyes associated with loss of horizontal eye movements and is diagnostic of bilateral pontine damage, usually from thrombosis of the basilar artery. “Ocular dipping” is a slower, arrhythmic downward movement followed by a faster upward movement in patients with normal reflex horizontal gaze; it usually indicates diffuse cortical anoxic damage.

The oculocephalic reflexes, elicited by moving the head from side to side or vertically and observing eye movements in the direction opposite to the head movement, depend on the integrity of the ocular motor nuclei and their interconnecting tracts that extend from the midbrain to the pons and medulla (Fig. 28-3). The movements, called somewhat inaccurately “doll’s eyes,” are normally suppressed in the awake patient with intact frontal lobes. The ability to elicit them therefore reflects both reduced cortical influence on the brainstem and intact brainstem pathways. The opposite, an absence of reflex eye movements, usually signifies damage within the brainstem but can result from overdoses of certain drugs. In this circumstance, normal pupillary size and light reaction distinguishes most drug-induced comas from structural brainstem damage. Oculocephalic maneuvers should not be attempted in patients with neck trauma, as vigorous head movements can precipitate or worsen a spinal cord injury.

Thermal, or “caloric,” stimulation of the vestibular apparatus (oculovestibular response) provides a more intense stimulus for the oculocephalic reflex but provides essentially the same information. The test is performed by irrigating the external auditory canal with cold water in order to induce convection currents in the labyrinths. After a brief latency, the result is tonic deviation of both eyes to the side of cold-water irrigation. In comatose patients, nystagmus in the opposite direction may not occur. The acronym “COWS” has been used to remind generations of medical students of the direction of nystagmus—cold water opposite, warm water same—but since nystagmus is often absent in the opposite direction due to frontal lobe dysfunction in coma, this mnemonic does not often hold true.

The corneal reflex, elicited by touching the cornea with a wisp of cotton and observing bilateral lid closure, depends on the integrity of pontine pathways between the fifth (afferent) and both seventh (efferent) cranial nerves; it is a useful test of pontine function. CNS-depressant drugs diminish or eliminate the corneal responses soon after reflex eye movements are paralyzed but before the pupils become unreactive to light. The corneal response may be lost for a time on the side of an acute hemiplegia.

Respiratory Patterns

These are of less localizing value in comparison to other brainstem signs. Shallow, slow, but regular breathing suggests metabolic or drug-induced depression of the medullary respiratory centers. Cheyne-Stokes respiration in its typical cyclic form, ending with a brief apneic period, signifies bihemispheral damage or metabolic suppression and commonly accompanies light coma. Rapid, deep (Kussmaul) breathing usually implies metabolic acidosis but may also occur with pontomesencephalic lesions. Agonal gasps are the result of lower brainstem (medullary) damage and are recognized as the terminal respiratory pattern of severe brain damage. Other cyclic breathing patterns have been described but are of lesser significance.

FIGURE 28-3

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser

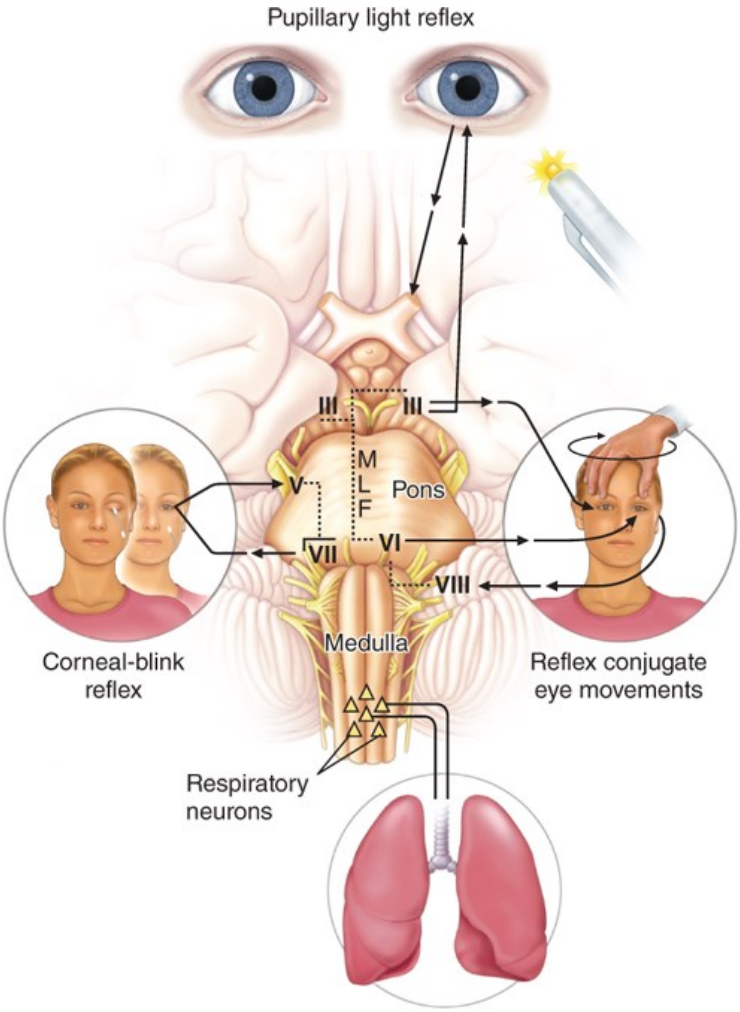
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

horizontal eye movements are dependent on the medial longitudinal fasciculus (MLF) interconnecting the sixth and contralateral third nerve nuclei. Head rotation (oculocephalic reflex) or caloric stimulation of the labyrinths (oculovestibular reflex) elicits contraversive eye movements (for details,

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

FIGURE 28-3

Examination of brainstem reflexes in coma. Midbrain and third nerve function are tested by pupillary reaction to light, pontine function by spontaneous and reflex eye movements and corneal responses, and medullary function by respiratory and pharyngeal responses. Reflex conjugate, horizontal eye movements are dependent on the medial longitudinal fasciculus (MLF) interconnecting the sixth and contralateral third nerve nuclei. Head rotation (oculocephalic reflex) or caloric stimulation of the labyrinths (oculovestibular reflex) elicits contraversive eye movements (for details, see text).



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

LABORATORY STUDIES AND IMAGING

The studies that are most useful in the diagnosis of coma are chemical-toxicologic analysis of blood and urine, cranial CT or MRI, EEG, and cerebrospinal fluid (CSF) examination. Arterial blood gas analysis is helpful in patients with lung disease and acid-base disorders. The metabolic aberrations commonly encountered in clinical practice are usually revealed by measurement of electrolytes, glucose, calcium, magnesium, osmolarity, and renal (blood urea nitrogen) and hepatic (NH_3) function. Toxicologic analysis may be necessary in cases of acute coma, when the diagnosis is not immediately clear. However, the presence of exogenous drugs or toxins, especially alcohol, does not exclude the possibility that other factors, particularly head trauma, are contributing to the clinical state. An ethanol level of 43 mmol/L (0.2 g/dL) in nonhabituated patients generally causes impaired mental activity; a level of >65 mmol/L (0.3 g/dL) is associated with stupor. The development of tolerance may allow some chronic alcoholics to remain awake at levels >87 mmol/L (0.4 g/dL).

The availability of cranial CT and MRI has focused attention on causes of coma that are detectable by imaging (e.g., hemorrhage, tumor, or hydrocephalus). Resorting primarily to this approach, although at times expedient, is imprudent because most cases of coma (and confusion) are

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

imaging results can be misleading such as when small subdural hematomas or old strokes are found, but the patient's coma is due to intoxication. Additional imaging with CT angiography or MRI can be obtained if acute posterior circulation stroke is considered.

The availability of cranial CT and MRI has focused attention on causes of coma that are detectable by imaging (e.g., hemorrhage, tumor, or hydrocephalus). Resorting primarily to this approach, although at times expedient, is imprudent because most cases of coma (and confusion) are metabolic or toxic in origin. Furthermore, a normal CT scan does not exclude an anatomic lesion as the cause of coma; for example, early bilateral hemisphere infarction, acute brainstem infarction, encephalitis, meningitis, mechanical shearing of axons as a result of closed head trauma, sagittal sinus thrombosis, hypoxic injury, and subdural hematoma isodense to adjacent brain are some of the disorders that may not be detected. Sometimes imaging results can be misleading such as when small subdural hematomas or old strokes are found, but the patient's coma is due to intoxication. Additional imaging with CT angiography or MRI can be obtained if acute posterior circulation stroke is considered.

The EEG **(Chap. 425)** provides clues in metabolic or drug-induced states but is rarely diagnostic in these disorders. However, it is the essential test to reveal coma due to nonconvulsive seizures and shows fairly characteristic patterns in herpesvirus encephalitis and prion disease. The EEG may be further helpful in disclosing generalized slowing of the background activity, a reflection of the severity of an encephalopathy. Predominant high-voltage slowing (δ or triphasic waves) in the frontal regions is typical of metabolic coma, as from hepatic failure, and widespread fast (β) activity implicates overdose with sedative drugs (e.g., benzodiazepines). A special pattern of "alpha coma," defined by widespread, variable 8- to 12-Hz activity, superficially resembles the normal α rhythm of waking but, unlike normal α activity, is not altered by environmental stimuli. Alpha coma results from pontine or diffuse cortical damage and is associated with a poor prognosis. A unique EEG pattern in adults of "extreme delta brush" is characteristic of a specific (anti–*N*-methyl-D-aspartate [NMDA] receptor) form of autoimmune encephalitis. Normal α activity on the EEG, which is suppressed by stimulating the patient, also alerts the clinician to the locked-in syndrome, hysteria, or catatonia.

Lumbar puncture should be performed if no cause is readily apparent, as examination of the CSF remains indispensable in the diagnosis of various forms of meningitis and encephalitis. An imaging study should be performed prior to lumbar puncture to exclude a large intracranial mass lesion, which could lead to herniation with lumbar puncture. Blood cultures and administration of antibiotics should precede the imaging study if infectious meningitis is suspected **(Chap. 138)**.

DIFFERENTIAL DIAGNOSIS OF COMA

(Table 28-1) The causes of coma can be divided into three broad categories: those without focal neurologic signs (e.g., metabolic and toxic encephalopathies); those with prominent focal signs (e.g., stroke, cerebral hemorrhage); and meningitis syndromes, characterized by fever or stiff neck and an excess of cells in the spinal fluid (e.g., bacterial meningitis, subarachnoid hemorrhage, encephalitis). Causes of sudden coma include drug ingestion, cerebral hemorrhage, trauma, cardiac arrest, epilepsy, and basilar artery occlusion. Coma that appears subacutely is usually related to a preexisting medical or neurologic problem or, less often, to secondary brain swelling surrounding a mass such as tumor or cerebral infarction.

TABLE 28-1

Differential Diagnosis of Coma

1. Diseases that cause no focal brainstem or lateralizing neurologic signs (CT scan is often normal)
   a. Intoxications: alcohol, sedative drugs, opiates, etc.
   b. Metabolic disturbances: anoxia, hyponatremia, hypernatremia, hypercalcemia, diabetic acidosis, nonketotic hyperosmolar hyperglycemia, hypoglycemia, uremia, hepatic coma, hypercarbia, Addisonian crisis, hypo- and hyperthyroid states, profound nutritional deficiency
   c. Severe systemic infections: pneumonia, septicemia, typhoid fever, malaria, Waterhouse-Friderichsen syndrome
   d. Shock from any cause
   e. Status epilepticus, nonconvulsive status epilepticus, postictal states
   f. Hyperperfusion syndromes including hypertensive encephalopathy, eclampsia, posterior reversible encephalopathy syndrome (PRES)
   g. Severe hyperthermia, hypothermia
   h. Concussion
   i. Acute hydrocephalus
2. Diseases that cause focal brainstem or lateralizing cerebral signs (CT scan is typically abnormal)
   a. Hemispheral hemorrhage (basal ganglionic, thalamic) or infarction (large middle cerebral artery territory) with secondary brainstem compression
   b. Brainstem infarction due to basilar artery thrombosis or embolism
   c. Brain abscess, subdural empyema
   d. Epidural and subdural hemorrhage, brain contusion
   e. Brain tumor with surrounding edema
3. Diseases that cause meningeal irritation with or without fever, and with an excess of white blood cells or red blood cells in the CSF
   a. Subarachnoid hemorrhage from ruptured aneurysm, arteriovenous malformation, trauma

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

preexisting medical or neurologic problem or, less often, to secondary brain swelling surrounding a mass such as tumor or cerebral infarction.

TABLE 28-1

Differential Diagnosis of Coma

1. Diseases that cause no focal brainstem or lateralizing neurologic signs (CT scan is often normal)
   a. Intoxications: alcohol, sedative drugs, opiates, etc.
   b. Metabolic disturbances: anoxia, hyponatremia, hypernatremia, hypercalcemia, diabetic acidosis, nonketotic hyperosmolar hyperglycemia, hypoglycemia, uremia, hepatic coma, hypercarbia, Addisonian crisis, hypo- and hyperthyroid states, profound nutritional deficiency
   c. Severe systemic infections: pneumonia, septicemia, typhoid fever, malaria, Waterhouse-Friderichsen syndrome
   d. Shock from any cause
   e. Status epilepticus, nonconvulsive status epilepticus, postictal states
   f. Hyperperfusion syndromes including hypertensive encephalopathy, eclampsia, posterior reversible encephalopathy syndrome (PRES)
   g. Severe hyperthermia, hypothermia
   h. Concussion
   i. Acute hydrocephalus
2. Diseases that cause focal brainstem or lateralizing cerebral signs (CT scan is typically abnormal)
   a. Hemispheral hemorrhage (basal ganglionic, thalamic) or infarction (large middle cerebral artery territory) with secondary brainstem compression
   b. Brainstem infarction due to basilar artery thrombosis or embolism
   c. Brain abscess, subdural empyema
   d. Epidural and subdural hemorrhage, brain contusion
   e. Brain tumor with surrounding edema
   f. Cerebellar and pontine hemorrhage and infarction
   g. Widespread traumatic brain injury
   h. Metabolic coma (see above) in the setting of preexisting focal damage
3. Diseases that cause meningeal irritation with or without fever, and with an excess of white blood cells or red blood cells in the CSF
   a. Subarachnoid hemorrhage from ruptured aneurysm, arteriovenous malformation, trauma
   b. Infectious meningitis and meningoencephalitis
   c. Paraneoplastic and autoimmune encephalitis
   d. Carcinomatous and lymphomatous meningitis

The diagnosis of coma due to cerebrovascular disease can be difficult **(Chap. 426)**. The most common diseases in this category are (1) basal ganglia and thalamic hemorrhage (acute but not instantaneous onset, vomiting, headache, hemiplegia, and characteristic eye signs); (2) pontine hemorrhage (sudden onset, pinpoint pupils, loss of reflex eye movements and corneal responses, ocular bobbing, posturing, and hyperventilation); (3) cerebellar hemorrhage (occipital headache, vomiting, gaze paresis, and inability to stand and walk); (4) basilar artery thrombosis (neurologic prodrome or transient ischemic attack warning spells, diplopia, dysarthria, vomiting, eye movement and corneal response abnormalities, and asymmetric limb paresis); and (5) subarachnoid hemorrhage (precipitous coma after sudden severe headache and vomiting). The most common stroke, infarction in the territory of the middle cerebral artery, does not cause coma, but edema surrounding large infarctions may expand over several days and cause coma from mass effect.

The syndrome of acute hydrocephalus accompanies many intracranial diseases, particularly subarachnoid hemorrhage. It is characterized by headache and sometimes vomiting that may progress quickly to coma with extensor posturing of the limbs, bilateral Babinski signs, small unreactive pupils, and impaired oculocephalic movements in the vertical direction. At times, the coma may be featureless without lateralizing signs, although papilledema is often present.

BRAIN DEATH

Brain death is a state of irreversible cessation of all cerebral and brainstem function with preservation of cardiac activity and maintenance of respiratory and somatic function by artificial means. It is the only type of brain damage recognized as morally, ethically, and legally equivalent to death. Criteria have been advanced for the diagnosis of brain death, and it is essential to adhere to consensus standards as multiple studies have shown

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Established criteria contain two essential elements, after assuring that no confounding factors (e.g., hypothermia, drug intoxication) are present: (1)

Criteria have been advanced for the diagnosis of brain death, and it is essential to adhere to consensus standards as multiple studies have shown variability in local practice. Given the implications of the diagnosis, clinicians must be thorough and precise in determining brain death. It is advisable to delay clinical testing for at least 24 h if a cardiac arrest has caused brain death or if the inciting disease is not known. Some centers advocate a brief period of observation between two examiners' tests during which the clinical signs of brain death are sustained.

Established criteria contain two essential elements, after assuring that no confounding factors (e.g., hypothermia, drug intoxication) are present: (1) widespread cortical destruction that is reflected by deep coma and unresponsiveness to all forms of stimulation; and (2) global brainstem damage as demonstrated by absent pupillary light reaction, absent corneal reflexes, loss of oculovestibular reflexes, and destruction of the medulla, manifested by complete and irreversible apnea. Diabetes insipidus is often present but may only develop hours or days after the other clinical signs of brain death appear. The pupils are usually midsized but may be enlarged. Loss of deep tendon reflexes is not required because the spinal cord remains functional. Occasionally, other reflexes that originate from the spine may be present and should not preclude a diagnosis of brain death.

Demonstration that apnea is due to medullary damage requires that the P_{CO_2} be high enough to stimulate respiration during a test of spontaneous breathing. Apnea testing can be done by the use of preoxygenation with 100% oxygen prior to and following removal of the ventilator. CO_2 tension increases ~0.3–0.4 kPa/min (2–3 mmHg/min) during apnea. Apnea is confirmed if no respiratory effort has been observed in the presence of a sufficiently elevated P_{CO_2}. The apnea test is usually stopped if there is cardiovascular instability and alternative means of testing can be employed.

An isoelectric EEG may be used as an optional confirmatory test for total cerebral damage. Radionuclide brain scanning, cerebral angiography, or transcranial Doppler measurements may be used to demonstrate the absence of blood flow when a confirmatory study is desired.

It is largely accepted in Western society that the ventilator can be disconnected from a brain-dead patient and that organ donation is subsequently possible. Good communication between the physician and the family is important with appropriate preparation of the family for brain death testing and diagnosis.

TREATMENT

Coma

The immediate goal in a comatose patient is prevention of further nervous system damage. Hypotension, hypoglycemia, hypercalcemia, hypoxia, hypercapnia, and hyperthermia should be corrected rapidly. Hyponatremia should be corrected slowly to avoid injury from osmotic demyelination **(Chap. 307)**. An oropharyngeal airway is adequate to keep the pharynx open in a drowsy patient who is breathing normally. Tracheal intubation is indicated if there is apnea, upper airway obstruction, hypoventilation, or emesis, or if the patient is at risk for aspiration. Mechanical ventilation is required if there is hypoventilation or a need to induce hypocapnia in order to lower ICP. **The management of raised ICP is discussed in Chap. 307.** IV access is established and naloxone and dextrose are administered if opioid overdose or hypoglycemia are possibilities; thiamine is given along with glucose to avoid provoking Wernicke's encephalopathy in malnourished patients. In cases of suspected ischemic stroke including basilar thrombosis with brainstem ischemia, IV tissue plasminogen activator or mechanical embolectomy is often used after cerebral hemorrhage has been excluded and when the patient presents within established time windows for these interventions **(Chap. 427)**. Physostigmine may awaken patients with anticholinergic-type drug overdose but should be used only with careful monitoring; many physicians believe that it should only be used to treat anticholinergic overdose–associated cardiac arrhythmias. The use of benzodiazepine antagonists offers some prospect of improvement after overdose; however, these drugs are not commonly used empirically in part due to their tendency to provoke seizures. Certain other toxic and drug-induced comas have specific treatments such as fomepizole for ethylene glycol ingestion.

Administration of hypotonic IV solutions should be monitored carefully in any serious acute brain illness because of the potential for exacerbating brain swelling. Cervical spine injuries must not be overlooked, particularly before attempting intubation or evaluation of oculocephalic responses. Fever and meningismus indicate an urgent need for examination of the CSF to diagnose meningitis. Whenever acute bacterial meningitis is suspected, antibiotics including at least vancomycin and a third-generation cephalosporin are typically administered rapidly along with dexamethasone **(see Chap. 138)**.

PROGNOSIS

Some patients, especially children and young adults, may have ominous early clinical findings such as abnormal brainstem reflexes and yet recover;

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Glasgow Coma Scale was devised; it has predictive value in cases of brain trauma **(see Chap. 443)**. For anoxic coma, clinical signs such as the pupillary and motor responses after 1 day, 3 days, and 1 week have predictive value; however, some prediction rules are less reliable in the setting of therapeutic

Some patients, especially children and young adults, may have ominous early clinical findings such as abnormal brainstem reflexes and yet recover; early prognostication outside of brain death therefore is unwise. Metabolic comas have a far better prognosis than traumatic ones. Systems for estimating prognosis in adults should be taken as approximations, and medical judgments must be tempered by factors such as age, underlying systemic disease, and general medical condition. In an attempt to collect prognostic information from large numbers of patients with head injury, the Glasgow Coma Scale was devised; it has predictive value in cases of brain trauma **(see Chap. 443)**. For anoxic coma, clinical signs such as the pupillary and motor responses after 1 day, 3 days, and 1 week have predictive value; however, some prediction rules are less reliable in the setting of therapeutic hypothermia, and therefore, serial examinations and multimodal prognostication approaches are advised in this setting. For example, the absence of the cortical responses of the somatosensory evoked potentials has been shown to be a strong indicator of poor outcome following hypoxic injury.

The poor outcome of persistent vegetative and minimally conscious states has already been mentioned, but reports of a small number of patients displaying cortical activation on functional MRI in response to salient stimuli have begun to alter the perception of such individuals. In one series, about 10% of vegetative patients (mainly following traumatic brain injury) could activate their frontal or temporal lobes in response to requests by an examiner to imagine certain visuospatial tasks. Another series demonstrated that up to 15% of patients with various forms of acute brain injury and absence of behavioral responses to motor commands showed EEG activation in response to these commands. It is prudent to avoid generalizations from these findings, but the need for future studies of novel techniques to help communication and possibly recovery is needed.

FURTHER READING

Claasen J et al: Detection of brain activation in unresponsive patients with acute brain injury. N Engl J Med 380:2497, 2019. [PubMed: 31242361]

Edlow JA et al: Diagnosis of reversible causes of coma. Lancet 384:2064, 2014. [PubMed: 24767707]

Greer DM et al: Determination of brain death/death by neurologic criteria: The World Brain Death Project. JAMA 324:1078, 2020. [PubMed: 32761206]

Monti MM et al: Willful modulation of brain activity in disorders of consciousness. N Engl J Med 362:579, 2010. [PubMed: 20130250]

Posner JB et al: *Plum and Posner's Diagnosis of Stupor and Coma*, 5th ed. New York, Oxford University Press, 2019.

Wijdicks EFM: Predicting the outcome of a comatose patient at the bedside. Pract Neurol 20:26, 2020. [PubMed: 31537635]

Downloaded 2022-1-3 5:46 P Your IP is 141.216.78.40
Chapter 28: Coma, S. Andrew Josephson; Allan H. Ropper; Stephen L. Hauser
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 29: Dementia

William W. Seeley; Gil D. Rabinovici; Bruce L. Miller

INTRODUCTION

Dementia, a syndrome with many causes, affects nearly 6 million people in the United States and results in a total annual health care cost in excess of $300 billion. Dementia is defined as an acquired deterioration in cognitive abilities that impairs the successful performance of activities of daily living. Episodic memory, the ability to recall events specific in time and place, is the cognitive function most commonly lost; 10% of persons age >70 years and 20–40% of individuals age >85 years have clinically identifiable memory loss. In addition to memory, dementia may erode other mental faculties, including language, visuospatial, praxis, calculation, judgment, and problem-solving abilities. Neuropsychiatric and social deficits also arise in many dementia syndromes, manifesting as depression, apathy, anxiety, hallucinations, delusions, agitation, insomnia, sleep disturbances, compulsions, or disinhibition. The clinical course may be slowly progressive, as in Alzheimer's disease (AD); static, as in anoxic encephalopathy; or may fluctuate from day to day or minute to minute, as in dementia with Lewy bodies (DLB). Most patients with AD, the most prevalent form of dementia, begin with episodic memory impairment, but in other dementias, such as frontotemporal dementia (FTD), memory loss is not typically a presenting feature. **Focal cerebral disorders are discussed in Chap. 30 and illustrated in a video library in Chap. V2; detailed discussions of AD can be found in Chap. 431; FTD and related disorders in Chap. 432; vascular dementia in Chap. 433; DLB in Chap. 434; Huntington's disease (HD) in Chap. 436; and prion diseases in Chap. 438.**

FUNCTIONAL ANATOMY OF THE DEMENTIAS

Dementia syndromes result from the disruption of specific large-scale neuronal networks; the location and severity of synaptic and neuronal loss combine to produce the clinical features (Chap. 30). Behavior, mood, and attention are modulated by ascending noradrenergic, serotonergic, and dopaminergic pathways, whereas cholinergic signaling is critical for attention and memory functions. The dementias differ in the relative neurotransmitter deficit profiles; accordingly, accurate diagnosis guides effective pharmacologic therapy.

AD typically begins in the entorhinal region of the medial temporal lobe, spreads to the hippocampus and other limbic structures, and moves through the basal temporal areas and then into the lateral and posterior temporal and parietal neocortex, eventually causing a more widespread degeneration. Vascular dementia is associated with focal damage in a variable patchwork of cortical and subcortical regions or white matter tracts that disconnects nodes within distributed networks. In keeping with its anatomy, AD typically presents with episodic memory loss accompanied later by aphasia, executive dysfunction, or navigational problems. In contrast, dementias that begin in frontal or subcortical regions, such as FTD or HD, are less likely to begin with memory problems and more likely to present with difficulties with judgment, mood, executive control, movement, and behavior.

Lesions of frontal-striatal[1] pathways produce specific and predictable effects on behavior. The dorsolateral prefrontal cortex has connections with a central band of the caudate nucleus. Lesions of either the caudate or dorsolateral prefrontal cortex, or their connecting white matter pathways, may result in executive dysfunction, manifesting as poor organization and planning, decreased cognitive flexibility, and impaired working memory. The lateral orbital frontal cortex connects with the ventromedial caudate, and lesions of this system cause impulsiveness, distractibility, and disinhibition. The anterior cingulate cortex and adjacent medial prefrontal cortex project to the nucleus accumbens, and interruption of this system produces apathy, poverty of speech, emotional blunting, or even akinetic mutism. All corticostriatal systems also include topographically organized projections through the globus pallidus and thalamus, and damage to these nodes can likewise reproduce the clinical syndrome associated with the corresponding cortical or striatal injuries. Involvement of brainstem nuclei and cerebellar structures can further contribute to cognitive, behavioral, and motor manifestations.

THE CAUSES OF DEMENTIA

The single strongest risk factor for dementia is increasing age. The prevalence of disabling memory loss increases with each decade over age 50 and is

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The many causes of dementia are listed in Table 29-1. The frequency of each condition depends on the age group under study, access of the group to

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

corresponding cortical or striatal injuries. Involvement of brainstem nuclei and cerebellar structures can further contribute to cognitive, behavioral

The single strongest risk factor for dementia is increasing age. The prevalence of disabling memory loss increases with each decade over age 50 and is usually associated with the microscopic changes of AD at autopsy. Yet some centenarians have intact memory function and no evidence of clinically significant dementia. Whether dementia is an inevitable consequence of normal human aging remains controversial although the prevalence increases with every decade of life.

The many causes of dementia are listed in **Table 29-1**. The frequency of each condition depends on the age group under study, access of the group to medical care, country of origin, and perhaps racial or ethnic background. AD is the most common cause of dementia in Western countries, accounting for more than half of all patients. Vascular disease is the second most frequent cause for dementia and is particularly common in elderly patients or populations with limited access to medical care, where vascular risk factors are undertreated. Often, vascular brain injury is mixed with neurodegenerative disorders, particularly AD, making it difficult, even for the neuropathologist, to estimate the contribution of cerebrovascular disease to the cognitive disorder in an individual patient. Dementias associated with Parkinson's disease (PD) are common and may develop years after onset of a parkinsonian disorder, as seen with PD-related dementia (PDD), or they can occur concurrently with or preceding the motor syndrome, as in DLB. Limbic-predominant aging-related TDP-43 encephalopathy (LATE) is common after age 70 and has been linked to declining episodic memory function. Chronic traumatic encephalopathy (CTE), a unique disease found in individuals with a history of repetitive head impacts (e.g., professional athletes in collision or fighting sports, military veterans exposed to multiple blasts), presents with changes in cognition, mood, behavior, or motor function. Mixed pathology is common, especially in older individuals. In patients under the age of 65, FTD rivals AD as the most common cause of dementia. Chronic intoxications, including those resulting from alcohol and prescription drugs, are an important and often treatable cause of dementia. Other disorders listed in **Table 29-1** are uncommon but important because many are reversible. The classification of dementing illnesses into reversible and irreversible disorders is a useful approach to differential diagnosis. When effective treatments for the neurodegenerative conditions emerge, this dichotomy will become obsolete.

TABLE 29-1

Differential Diagnosis of Dementia

Most Common Causes of Dementia	
Alzheimer's disease	Alcoholism[a]
Vascular dementia	PDD/LBD spectrum
Multi-infarct	Drug/medication intoxication[a]
Diffuse white matter disease (Binswanger's)	Limbic-predominant age-related TDP-43 encephalopathy
Less Common Causes of Dementia	
Vitamin deficiencies Thiamine (B_1): Wernicke's encephalopathy[a] B_{12} (subacute combined degeneration)[a] Nicotinic acid (pellagra)[a] Endocrine and other organ failure Hypothyroidism[a] Adrenal insufficiency and Cushing's syndrome[a] Hypo- and hyperparathyroidism[a] Renal failure[a] Liver failure[a] Pulmonary failure[a] Chronic infections	Toxic disorders Drug, medication, and narcotic poisoning[a] Heavy metal intoxication[a] Organic toxins Psychiatric Depression (pseudodementia)[a] Schizophrenia[a] Conversion disorder[a] Degenerative disorders Huntington's disease Multisystem atrophy Hereditary ataxias (some forms) Frontotemporal lobar degeneration spectrum
leukoencephalopathy) Tuberculosis, fungal, and protozoal[a]	Prion (Creutzfeldt-Jakob and Gerstmann-Straussler-Scheinker diseases) Miscellaneous

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Most Common Causes of Dementia	
Alzheimer's disease	Alcoholism[a]
Vascular dementia	PDD/LBD spectrum
Multi-infarct	Drug/medication intoxication[a]
Diffuse white matter disease (Binswanger's)	Limbic-predominant age-related TDP-43 encephalopathy
Less Common Causes of Dementia	
Vitamin deficiencies Thiamine (B_1): Wernicke's encephalopathy[a] B_{12} (subacute combined degeneration)[a] Nicotinic acid (pellagra)[a] Endocrine and other organ failure Hypothyroidism[a] Adrenal insufficiency and Cushing's syndrome[a] Hypo- and hyperparathyroidism[a] Renal failure[a] Liver failure[a] Pulmonary failure[a] Chronic infections HIV Neurosyphilis[a] Papovavirus (JC virus) (progressive multifocal leukoencephalopathy) Tuberculosis, fungal, and protozoal[a] Whipple's disease[a] Head trauma and diffuse brain damage Chronic traumatic encephalopathy Chronic subdural hematoma[a] Postanoxia Postencephalitis Normal-pressure hydrocephalus[a] Intracranial hypotension Neoplastic Primary brain tumor[a] Metastatic brain tumor[a] Autoimmune (paraneoplastic) encephalitis[a]	Toxic disorders Drug, medication, and narcotic poisoning[a] Heavy metal intoxication[a] Organic toxins Psychiatric Depression (pseudodementia)[a] Schizophrenia[a] Conversion disorder[a] Degenerative disorders Huntington's disease Multisystem atrophy Hereditary ataxias (some forms) Frontotemporal lobar degeneration spectrum Multiple sclerosis Adult Down's syndrome with Alzheimer's disease ALS-parkinsonism-dementia complex of Guam Prion (Creutzfeldt-Jakob and Gerstmann-Sträussler-Scheinker diseases) Miscellaneous Sarcoidosis[a] Vasculitis[a] CADASIL, etc. Acute intermittent porphyria[a] Recurrent nonconvulsive seizures[a] Additional conditions in children or adolescents Pantothenate kinase–associated neurodegeneration Subacute sclerosing panencephalitis Metabolic disorders (e.g., Wilson's and Leigh's diseases, leukodystrophies, lipid storage diseases, mitochondrial mutations)

[a] Potentially reversible dementia.

Abbreviations: ALS, amyotrophic lateral sclerosis; CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; LBD,

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

potentially reversible concomitant condition that may have contributed to the patient's impairment. The three most common potentially reversible diagnoses were depression, normal pressure hydrocephalus (NPH), and alcohol dependence; medication side effects are also common and should be

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Autoimmune (paraneoplastic) encephalitis[a]

Abbreviations: ALS, amyotrophic lateral sclerosis; CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; LBD, Lewy body disease; PDD, Parkinson's disease dementia.

In a study of 1000 persons attending a memory disorders clinic, 19% had a potentially reversible cause of the cognitive impairment and 23% had a potentially reversible concomitant condition that may have contributed to the patient's impairment. The three most common potentially reversible diagnoses were depression, normal pressure hydrocephalus (NPH), and alcohol dependence; medication side effects are also common and should be considered in every patient **(Table 29-1)**.

The term *rapidly progressive dementia (RPD)* is applied to illnesses that progress from initial symptom onset to dementia within a year or less; confusional states related to toxic/metabolic conditions are excluded. Although the prion proteinopathy Creutzfeldt-Jakob disease (CJD) **(Chap. 438)** is the classic cause of a rapidly progressive dementia, especially when associated with myoclonus, more often cases of RPD are due to AD or another neurodegenerative disorder, or to an autoimmune encephalitis.

Subtle cumulative decline in episodic memory is a common part of aging. This frustrating experience, often the source of jokes and humor, has historically been referred to as *benign forgetfulness of the elderly. Benign* means that it is not so progressive or serious that it impairs successful and productive daily functioning, although the distinction between benign and significant memory loss can be subtle. At age 85, the average person is able to learn and recall approximately one-half of the items (e.g., words on a list) that he or she could at age 18. The term *subjective cognitive decline* describes individuals who experience a subjective decline from their cognitive baseline but perform within normal limits for their age and educational attainment on formal neuropsychological testing. *Mild cognitive impairment (MCI)* is defined as a decline in cognition that is confirmed on objective cognitive testing but does not disrupt normal daily activities. MCI can be further subcategorized based on the presenting complaints and deficits (e.g., amnestic MCI, executive MCI). Factors that predict progression from MCI to an AD dementia include a prominent memory deficit, family history of dementia, presence of an apolipoprotein ε4 (Apo ε4) allele, small hippocampal volumes, an AD-like signature of cortical atrophy, low cerebrospinal fluid Aβ and elevated tau, or evidence of brain amyloid and tau deposition on positron emission tomography (PET) imaging.

The major degenerative dementias include AD, DLB, FTD and related disorders, HD, and prion diseases, including CJD. All are associated with the abnormal aggregation of a specific protein: $A\beta_{42}$ and tau in AD; α-synuclein in DLB; tau, TAR DNA-binding protein of 43 kDa (TDP-43), or the FET family of proteins (*fu*sed in *s*arcoma [FUS], Ewing sarcoma [EWS], and TBP-associated factor 15 [TAF15]) in FTD; huntingtin in HD; and misfolded prion protein (PrP^{Sc}) in CJD **(Table 29-2)**.

TABLE 29-2

The Molecular Basis for Degenerative Dementia

DEMENTIA	MOLECULAR BASIS	CAUSAL GENES (CHROMOSOME)	SUSCEPTIBILITY GENES	PATHOLOGIC FINDINGS
AD	Aβ/tau	*APP* (21), *PS-1* (14), *PS-2* (1) (<2% carry these mutations, most often in *PS-1*)	*Apo ε4* (19)	Amyloid plaques, neurofibrillary tangles, and neuropil threads
FTD	Tau	*MAPT* exon and intron mutations (17) (about 10% of familial cases)	H1 *MAPT* haplotype	Tau neuronal and glial inclusions varying in morphology and distribution
	TDP-43	*GRN* (10% of familial cases), *C9ORF72* (20%–30% of familial cases), rare *VCP*, very rare *TARDBP, TBK1, TIA1*		TDP-43 neuronal and glial inclusions varying in morphology and distribution
	FET	Very rare *FUS*		FET neuronal and glial inclusions varying in morphology and distribution
DLB	α-Synuclein	Very rare *SNCA* (4)	Unknown	α-Synuclein neuronal inclusions
		dominant mutations)	homozygosity for methionine or valine	spongiosis

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 29-2

The Molecular Basis for Degenerative Dementia

DEMENTIA	MOLECULAR BASIS	CAUSAL GENES (CHROMOSOME)	SUSCEPTIBILITY GENES	PATHOLOGIC FINDINGS
AD	Aβ/tau	*APP* (21), *PS-1* (14), *PS-2* (1) (<2% carry these mutations, most often in *PS-1*)	*Apo ε4* (19)	Amyloid plaques, neurofibrillary tangles, and neuropil threads
FTD	Tau	*MAPT* exon and intron mutations (17) (about 10% of familial cases)	H1 *MAPT* haplotype	Tau neuronal and glial inclusions varying in morphology and distribution
	TDP-43	*GRN* (10% of familial cases), *C9ORF72* (20%–30% of familial cases), rare *VCP*, very rare *TARDBP, TBK1, TIA1*		TDP-43 neuronal and glial inclusions varying in morphology and distribution
	FET	Very rare *FUS*		FET neuronal and glial inclusions varying in morphology and distribution
DLB	α-Synuclein	Very rare *SNCA* (4)	Unknown	α-Synuclein neuronal inclusions (Lewy bodies)
CJD	PrPSC	*PRNP* (20) (up to 15% of patients carry these dominant mutations)	Codon 129 homozygosity for methionine or valine	PrPSC deposition, panlaminar spongiosis

Abbreviations: AD, Alzheimer's disease; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FET, FUS/EWS/TAF-15; FTD, frontotemporal dementia.

The risk of developing dementia in late-life is associated with exposures and lifestyle factors that can operate across the life span. Modifiable risk factors include low education, hearing loss, traumatic brain injury, hypertension, diabetes mellitus, obesity, heavy alcohol use, smoking, depression, physical inactivity, and air pollution. Improved management of midlife vascular risk factors has been credited with a decreasing incidence of dementia observed in North America and Western Europe.

APPROACH TO THE PATIENT

Dementias

Three major issues should be kept at the forefront: (1) What is the clinical diagnosis? (2) What component of the dementia syndrome is treatable or reversible? (3) Can the physician help to alleviate the burden on caregivers? A broad overview of the approach to dementia is shown in **Table 29-3**. The major degenerative dementias can usually be distinguished by the initial symptoms; neuropsychological, neuropsychiatric, and neurologic findings; and neuroimaging features **(Table 29-4)**.

HISTORY

The history should concentrate on the onset, duration, and tempo of progression. An acute or subacute onset of confusion may be due to delirium **(Chap. 27)** and should trigger a search for intoxication, infection, or metabolic derangement. An elderly person with slowly progressive memory loss over several years is likely to suffer from AD. Nearly 75% of patients with AD begin with memory symptoms, but other early symptoms include anxiety or depression as well as difficulty managing money, driving, shopping, following instructions, finding words, or navigating. Personality change, disinhibition, and weight gain or compulsive eating suggest FTD, not AD. FTD is also suggested by prominent apathy, compulsivity, loss of

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

syndrome, the delusion that a familiar person has been replaced by an impostor.

A history of stroke with irregular stepwise progression suggests vascular dementia. Vascular dementia is also commonly seen in the setting of

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

change, disinhibition, and weight gain or compulsive eating suggest FTD, not AD. FTD is also suggested by prominent apathy, compulsivity, loss of empathy for others, or progressive loss of speech fluency or single-word comprehension with relative sparing of memory and visuospatial abilities. The diagnosis of DLB is suggested by early visual hallucinations; parkinsonism; proneness to delirium or sensitivity to psychoactive medications; rapid eye movement (REM) behavior disorder (RBD; dramatic, sometimes violent, limb movements during dreaming **[Chap. 31]**); or Capgras syndrome, the delusion that a familiar person has been replaced by an impostor.

A history of stroke with irregular stepwise progression suggests vascular dementia. Vascular dementia is also commonly seen in the setting of hypertension, atrial fibrillation, peripheral vascular disease, smoking, and diabetes. In patients suffering from cerebrovascular disease, it can be difficult to determine whether the dementia is due to AD, vascular disease, or a mixture of the two because many of the risk factors for vascular dementia, including diabetes, high cholesterol, elevated homocysteine, and low exercise, are also risk factors for AD. Moreover, many patients with a major vascular contribution to their dementia lack a history of stepwise decline. Rapid progression with motor rigidity and myoclonus suggests CJD **(Chap. 438)**. Seizures may indicate strokes or neoplasm but also occur in AD, particularly early-age-of-onset AD. Gait disturbance is common in vascular dementia, PD/DLB, or NPH. A history of high-risk sexual behaviors or intravenous drug use should trigger a search for central nervous system (CNS) infection, especially HIV or syphilis. A history of recurrent head trauma could indicate chronic subdural hematoma, CTE, intracranial hypotension, or NPH. Subacute onset of severe amnesia and psychosis with mesial temporal T2/fluid-attenuated inversion recovery (FLAIR) hyperintensities on MRI should raise concern for autoimmune (paraneoplastic) encephalitis, sometimes in long-term smokers or other patients at risk for cancer. The spectrum of autoimmune etiologies producing RPD has rapidly expanded, and includes antibodies targeting leucine-rich glioma-inactivated 1 (LGI1; faciobrachial dystonic seizures); contactin-associated protein-like 2 (Caspr2; insomnia, ataxia, myotonia); *N-methyl-D-aspartate* (NMDA)-receptor (psychosis, insomnia, dyskinesias); and α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA)-receptor (limbic encephalitis with relapses), among others **(Chap. 94)**. Alcohol abuse creates risk for malnutrition and thiamine deficiency. Veganism, bowel irradiation, an autoimmune diathesis, a remote history of gastric surgery, and chronic therapy with histamine H2-receptor antagonists for dyspepsia or gastroesophageal reflux predispose to B_{12} deficiency. Certain occupations, such as working in a battery or chemical factory, might indicate heavy metal intoxication. Careful review of medication intake, especially for sedatives and analgesics, may raise the issue of chronic drug intoxication. An autosomal dominant family history is found in HD and in familial forms of AD, FTD, DLB, or prion disorders. A history of mood disorder, the recent death of a loved one, or depressive signs such as insomnia or weight loss, raise the possibility of depression-related cognitive impairment.

PHYSICAL AND NEUROLOGIC EXAMINATION

A thorough general and neurologic examination is essential to identify signs of nervous system involvement and search for clues suggesting a systemic disease that might be responsible for the cognitive disorder. Typical AD spares motor systems until late in the course. In contrast, patients with FTD often develop axial rigidity, supranuclear gaze palsy, or a motor neuron disease reminiscent of amyotrophic lateral sclerosis (ALS). In DLB, the initial symptoms may include a parkinsonian syndrome (resting tremor, cogwheel rigidity, bradykinesia, festinating gait), but DLB often starts with visual hallucinations or cognitive impairment, and symptoms referable to the lower brainstem (RBD, gastrointestinal, or autonomic problems) may arise years or even decades before parkinsonism or dementia. Corticobasal syndrome (CBS) features asymmetric akinesia and rigidity, dystonia, myoclonus, alien limb phenomena, pyramidal signs, and prefrontal deficits such as nonfluent aphasia with or without motor speech impairment, executive dysfunction, apraxia, or a behavioral disorder. Progressive supranuclear palsy (PSP) is associated with unexplained falls, axial rigidity, dysphagia, and vertical gaze deficits. CJD is suggested by the presence of diffuse rigidity, an akinetic mute state, and prominent, often startle-sensitive, myoclonus.

Hemiparesis or other focal neurologic deficits suggest vascular dementia or brain tumor. Dementia with a myelopathy and peripheral neuropathy suggests vitamin B_{12} deficiency. Peripheral neuropathy could also indicate another vitamin deficiency, heavy metal intoxication, thyroid dysfunction, Lyme disease, or vasculitis. Dry cool skin, hair loss, and bradycardia suggest hypothyroidism. Fluctuating confusion associated with repetitive stereotyped movements may indicate ongoing limbic, temporal, or frontal seizures. In the elderly, hearing impairment or visual loss may produce confusion and disorientation misinterpreted as dementia. Profound bilateral sensorineural hearing loss in a younger patient with short stature or myopathy, however, should raise concern for a mitochondrial disorder.

COGNITIVE AND NEUROPSYCHIATRIC EXAMINATION

Brief screening tools such as the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MOCA), the Tablet Based Cognitive Assessment Tool, and Cognistat can be used to capture dementia and follow progression. None of these tests is highly sensitive to early-stage dementia or reliably discriminates between dementia syndromes. The MMSE is a 30-point test of cognitive function, with each correct answer being scored as 1 point. It includes tests of: orientation (e.g., identify season/date/month/year/floor/hospital/town/state/country); registration (e.g., name and restate 3 objects); recall (e.g., remember the same three objects 5 minutes later); and language (e.g., name pencil and watch; repeat “no

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

tasks of working and episodic memory, executive function, language, and visuospatial and perceptual abilities. In AD, the early deficits involve episodic memory, category generation (“name as many animals as you can in 1 minute”), and visuoconstructive ability. Usually deficits in verbal or

name and restate 3 objects); recall (e.g., remember the same three objects 5 minutes later); and language (e.g., name pencil and watch; repeat "no ifs ands or buts"; follow a 3-step command; obey a written command; and write a sentence and copy a design). In most patients with MCI and some with clinically apparent AD, bedside screening tests may be normal, and a more challenging and comprehensive set of neuropsychological tests will be required. When the etiology for the dementia syndrome remains in doubt, a specially tailored evaluation should be performed that includes tasks of working and episodic memory, executive function, language, and visuospatial and perceptual abilities. In AD, the early deficits involve episodic memory, category generation ("name as many animals as you can in 1 minute"), and visuoconstructive ability. Usually deficits in verbal or visual episodic memory are the first neuropsychological abnormalities detected, and tasks that require the patient to recall a long list of words or a series of pictures after a predetermined delay will demonstrate deficits in most patients. In FTD, the earliest deficits on cognitive testing involve executive control or language (speech or naming) functions, but some patients lack either finding despite profound social-emotional deficits. PDD or DLB patients have more severe deficits in executive and visuospatial function but do better on episodic memory tasks than patients with AD. Patients with vascular dementia often demonstrate a mixture of executive and visuospatial deficits, with prominent psychomotor slowing. In delirium, the most prominent deficits involve attention, working memory, and executive function, making the assessment of other cognitive domains challenging and often uninformative.

A functional assessment should also be performed to help the physician determine the day-to-day impact of the disorder on the patient's memory, community affairs, hobbies, judgment, dressing, and eating. Knowledge of the patient's functional abilities will help the clinician and the family to organize a therapeutic approach.

Neuropsychiatric assessment is important for diagnosis, prognosis, and treatment. In the early stages of AD, mild depressive features, social withdrawal, and irritability or anxiety are the most prominent psychiatric changes, but patients often maintain core social graces into the middle or late stages, when delusions, agitation, and sleep disturbance may emerge. In FTD, dramatic personality change with apathy, overeating, compulsions, disinhibition, and loss of empathy are early and common. DLB is associated with visual hallucinations, delusions related to person or place identity, RBD, and excessive daytime sleepiness. Dramatic fluctuations occur not only in cognition but also in arousal. Vascular dementia can present with psychiatric symptoms such as depression, anxiety, delusions, disinhibition, or apathy.

LABORATORY TESTS

The choice of laboratory tests in the evaluation of dementia is complex and should be tailored to the individual patient. The physician must take measures to avoid missing a reversible or treatable cause, yet no single treatable etiology is common; thus a screen must use multiple tests, each of which has a low yield. Cost/benefit ratios are difficult to assess, and many laboratory screening algorithms for dementia discourage multiple tests. Nevertheless, even a test with only a 1–2% positive rate is worth undertaking if the alternative is missing a treatable cause of dementia. **Table 29-3** lists most screening tests for dementia. The American Academy of Neurology recommends the routine measurement of a complete blood count; electrolytes; glucose; renal, liver, and thyroid functions; a vitamin B_{12} level; and a structural neuroimaging study (MRI or CT).

Neuroimaging studies, especially MRI, help to rule out primary and metastatic neoplasms, locate areas of infarction or inflammation, detect subdural hematomas, and suggest NPH or diffuse white matter disease. They also help to establish a regional pattern of atrophy. Support for the diagnosis of AD includes hippocampal atrophy in addition to posterior-predominant cortical atrophy **(Fig. 29-1)**. Focal frontal, insular, and/or anterior temporal atrophy suggests FTD **(Chap. 432)**. DLB often features less prominent atrophy, with greater involvement of the amygdala than the hippocampus. In CJD, magnetic resonance (MR) diffusion-weighted imaging reveals restricted diffusion within the cortical ribbon and/or basal ganglia in most patients. Extensive multifocal white matter abnormalities suggest a vascular etiology **(Fig. 29-2)**. Communicating hydrocephalus with vertex effacement (crowding of dorsal convexity gyri/sulci), gaping Sylvian fissures despite minimal cortical atrophy, and additional features shown in **Fig. 29-3** suggest NPH. Single-photon emission computed tomography (SPECT) and fluoro-deoxyglucose PET scanning show temporal-parietal hypoperfusion or hypometabolism in AD and frontotemporal deficits in FTD, but abnormalities in these patterns can be detected with MRI alone in many patients. Recently, amyloid- and tau-PET imaging have shown promise for the diagnosis of AD. There are currently three amyloid PET ligands (F18-florbetapir, F18-florbetaben, F18-flutametamol) and one tau PET ligand (F18-flortaucipir) approved by the US Food and Drug Administration for clinical use. Amyloid PET ligands bind to diffuse and neuritic amyloid plaques, as well as to vascular amyloid deposits (prominent in cerebral amyloid angiopathy), while tau PET ligands bind to the paired helical filaments of tau characteristic of neurofibrillary tangles in AD **(Chap. 431)**. Because amyloid plaques are also commonly found in cognitively normal older persons (~25% of individuals at age 65), the main clinical value of amyloid imaging is to exclude AD as the likely cause of dementia in patients who have negative scans. The spread of tau is more tightly linked to cognitive state **(Chap. 431)**, and thus may be more useful than amyloid imaging for "ruling in" AD, as well as for disease staging. Once disease-modifying therapies become available, CSF or molecular PET biomarkers will likely be used to identify treatment candidates. In the meantime, the prognostic value of detecting brain amyloid in an asymptomatic elder to assess preclinical disease and risk of future cognitive decline remains a topic of vigorous investigation.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

presence of low $A\beta_{42}$ (or a low $A\beta_{42}/A\beta_{40}$ ratio), mild-moderately elevated CSF total tau, and elevated CSF phosphorylated tau (at residues 181 or

ACCESS Medicine®

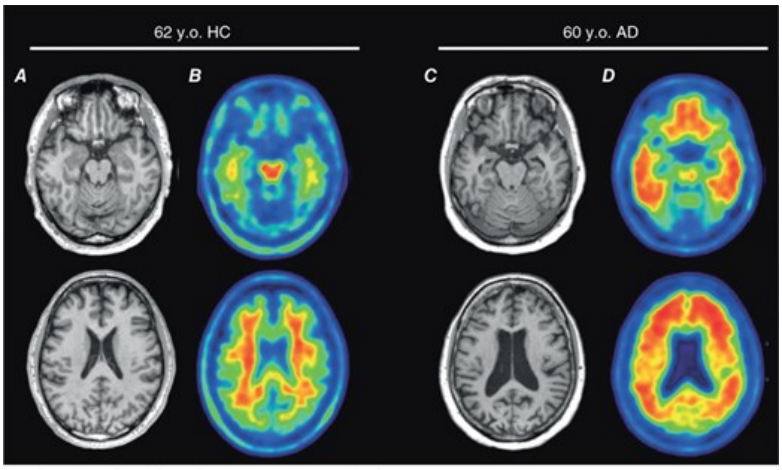
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

meantime, the prognostic value of detecting brain amyloid in an asymptomatic elder to assess preclinical disease and risk of future cognitive decline remains a topic of vigorous investigation.

Lumbar puncture need not be done routinely in the evaluation of dementia, but it is indicated when CNS infection or inflammation are credible diagnostic possibilities. Cerebrospinal fluid (CSF) levels of $A\beta_{42}$ and tau proteins show differing patterns with the various dementias, and the presence of low $A\beta_{42}$ (or a low $A\beta_{42}/A\beta_{40}$ ratio), mild-moderately elevated CSF total tau, and elevated CSF phosphorylated tau (at residues 181 or 217) is highly suggestive of AD. Novel fully automated CSF Aβ and tau assays perform comparably to amyloid and tau PET respectively, though, as with PET, their routine use in the diagnosis of dementia is debated. Blood-based biomarkers for AD show promise as a less invasive screening tool but remain under development **(Chap. 431)**. Formal psychometric testing helps to document the severity of cognitive disturbance, suggests psychogenic causes, and provides a more formal method for following the disease course. Electroencephalogram (EEG) is not routinely used but can help to suggest CJD (repetitive bursts of diffuse high-amplitude sharp waves, or "periodic complexes") or an underlying nonconvulsive seizure disorder (epileptiform discharges). Brain biopsy (including meninges) is not advised except to diagnose vasculitis, neoplasms, or unusual infections when the diagnosis is uncertain. Systemic disorders with CNS manifestations, such as sarcoidosis, can often be confirmed through biopsy of lymph node or solid organ rather than brain. MR angiography should be considered when cerebral vasculitis or cerebral venous thrombosis is a possible cause of the dementia.

FIGURE 29-1

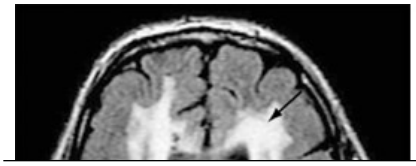
Alzheimer's disease (AD). Axial T1-weighted magnetic resonance images of a healthy 62-year-old (***A, B***) and a 60-year-old with AD (***C, D***). Note the diffuse atrophy, plus temporal lobe volume loss, in the patient with AD. Aβ positron emission tomography (PET) with [^{11}C]PIB (***B*** and ***D***) reveals extensive radiotracer retention in neocortex bilaterally in AD, consistent with the known distribution of amyloid plaques. HC, healthy control. *(Source: Gil Rabinovici, University of California, San Francisco and William Jagust, University of California, Berkeley.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 29-2

Diffuse white matter disease. Axial fluid-attenuated inversion recovery (FLAIR) magnetic resonance image through the lateral ventricles reveals multiple areas of hyperintensity (*arrows*) involving the periventricular white matter as well as the corona radiata and striatum. Although seen in some individuals with normal cognition, this appearance is more pronounced in patients with dementia of a vascular etiology. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

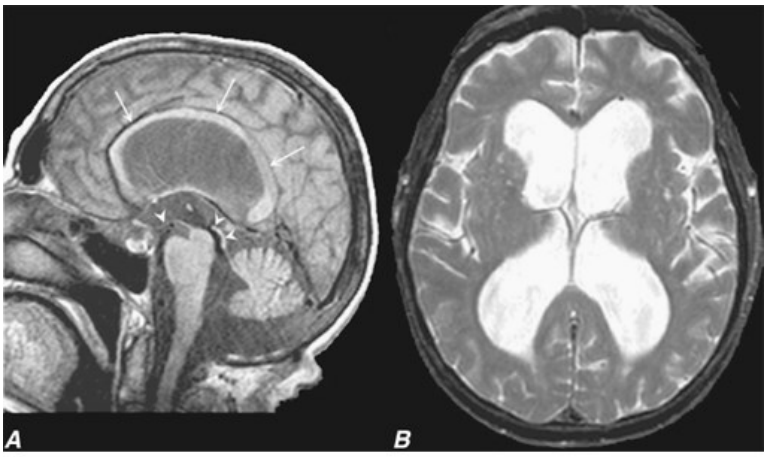
ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 29-3

Normal pressure hydrocephalus. ***A.*** Sagittal T1-weighted MRI demonstrates dilation of the lateral ventricle and stretching of the corpus callosum (*arrows*), depression of the floor of the third ventricle (*single arrowhead*), and enlargement of the aqueduct (*double arrowheads*). Note the diffuse dilation of the lateral, third, and fourth ventricles with a patent aqueduct, typical of communicating hydrocephalus. ***B.*** Axial T2-weighted MRIs demonstrate dilation of the lateral ventricles. This patient underwent successful ventriculoperitoneal shunting.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 29-3

Evaluation of the Patient with Dementia

ROUTINE EVALUATION	OPTIONAL FOCUSED TESTS	OCCASIONALLY HELPFUL TESTS
History	Psychometric testing	EEG
Physical examination	Chest x-ray	Parathyroid function
Laboratory tests	Lumbar puncture	Adrenal function
Thyroid function (TSH)	Liver function	Urine heavy metals
Vitamin B12	Renal function	RBC sedimentation rate
Electrolytes	HIV	Brain biopsy

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Vitamin B_{12}	Renal function	RBC sedimentation rate
Complete blood count	Urine toxin screen	Angiogram
Electrolytes	HIV	Brain biopsy
CT/MRI	Apolipoprotein E	SPECT
	RPR or VDRL	PET
		Autoantibodies
Diagnostic Categories		
REVERSIBLE CAUSES	**IRREVERSIBLE/DEGENERATIVE DEMENTIAS**	**PSYCHIATRIC DISORDERS**
Examples	Examples	Depression
Hypothyroidism	Alzheimer's	Schizophrenia
Thiamine deficiency	Frontotemporal dementia	Conversion reaction
Vitamin B_{12} deficiency	Huntington's	
Normal pressure hydrocephalus	Dementia with Lewy bodies	
Subdural hematoma	Vascular	
Chronic infection	Leukoencephalopathies	
Brain tumor	Parkinson's	
Drug intoxication		
Autoimmune encephalopathy		
Associated Treatable Conditions		
	Depression	Agitation
	Seizures	Caregiver "burnout"
	Insomnia	Drug side effects

Abbreviations: CT, computed tomography; EEG, electroencephalogram; MRI, magnetic resonance imaging; PET, positron emission tomography; RBC, red blood cell; RPR, rapid plasma reagin (test); SPECT, single-photon emission computed tomography; TSH, thyroid-stimulating hormone; VDRL, venereal disease research laboratory (test for syphilis).

TABLE 29-4

Clinical Differentiation of the Major Dementias

		loss	depression		hippocampal atrophy

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

laboratory (test for syphilis).

TABLE 29-4

Clinical Differentiation of the Major Dementias

DISEASE	FIRST SYMPTOM	MENTAL STATUS	NEUROPSYCHIATRY	NEUROLOGY	IMAGING
AD	Memory loss	Episodic memory loss	Irritability, anxiety, depression	Initially normal	Entorhinal cortex and hippocampal atrophy
FTD	Apathy, poor judgment/insight, speech/language, hyperorality	Frontal/executive and/or language; spares drawing	Apathy, disinhibition, overeating, compulsivity	May have vertical gaze palsy, axial rigidity, dystonia, alien hand, or MND	Frontal, insular, and/or temporal atrophy; usually spares posterior parietal lobe
DLB	Visual hallucinations, REM sleep behavior disorder, delirium, Capgras syndrome, parkinsonism	Drawing and frontal/executive, spares memory, delirium-prone	Visual hallucinations, depression, sleep disorder, delusions	Parkinsonism	Posterior parietal atrophy, hippocampi larger than in AD
CJD	Dementia, mood, anxiety, movement disorders	Variable, frontal/executive, focal cortical, memory	Depression, anxiety, psychosis in some	Myoclonus, rigidity, parkinsonism	Cortical ribboning and basal ganglia or thalamus hyperintensity on diffusion/FLAIR MRI
Vascular	Often but not always sudden, variable, apathy, falls, focal weakness	Frontal/executive, cognitive slowing, can spare memory	Apathy, delusions, anxiety	Usually motor slowing, spasticity, can be normal	Cortical and/or subcortical infarctions, confluent white matter disease

Abbreviations: AD, Alzheimer's disease; CBD, cortical basal degeneration; CJD, Creutzfeldt-Jakob disease; DLB, dementia with Lewy bodies; FLAIR, fluid-attenuated inversion recovery; FTD, frontotemporal dementia; MND, motor neuron disease; MRI, magnetic resonance imaging; REM, rapid eye movement.

[1]The striatum comprises the caudate/putamen/nucleus accumbens.

GLOBAL CONSIDERATIONS

Vascular dementia **(Chap. 433)** is more common in Asia due to the higher prevalence of intracranial atherosclerosis. Rates of vascular dementia are also on the rise in developing countries as vascular risk factors such as hypertension, hypercholesterolemia, and diabetes mellitus become more widespread. CNS infections, HIV (and associated opportunistic infections), syphilis, cysticercosis, and tuberculosis, likewise represent major contributors to dementia in the developing world. Systemic infection with SARS-CoV-2 may, in some individuals, have lasting effects on cognition due to involvement of brain microvasculature or to immunologically mediated white matter injury (acute disseminated encephalomyelitis [ADEM]) **(Chap. 444)**. Some individuals complain of lasting fatigue, changes in mood, and cognitive difficulties, but the long-term prognosis for SARS-CoV-2-related cognitive impairment remains unknown. Isolated populations have also contributed to our understanding of neurodegenerative dementia. Kuru, the cannibalism-associated rapidly progressive dementia seen in tribal New Guinea, played a role in the discovery of human prion disease. Amyotrophic lateral sclerosis-parkinsonism-dementia complex of Guam (or, Lytico-bodig disease) is a poly-proteinopathy, often with tau, TDP-43, and alpha-synuclein aggregation. The root cause of the disease remains uncertain, but its incidence has declined sharply over the past 60 years.

TREATMENT

Dementia

The major goals of dementia management are to treat reversible causes and provide comfort and support to the patient and caregivers. Treatment

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

patient's cognitive complaints stem from a psychiatric disorder, vigorous treatment of the condition should be tried to eliminate the cognitive complaint or to confirm that it persists despite adequate resolution of the mood or anxiety symptoms. Patients with degenerative diseases may also

The major goals of dementia management are to treat reversible causes and provide comfort and support to the patient and caregivers. Treatment of underlying causes includes thyroid replacement for hypothyroidism; vitamin therapy for thiamine or B_{12} deficiency or for elevated serum homocysteine; antimicrobials for opportunistic infections or antiretrovirals for HIV; ventricular shunting for NPH; or surgical, radiation, and/or chemotherapeutic treatment for CNS neoplasms. Removal of cognition-impairing drugs or medications is essential when appropriate. If the patient's cognitive complaints stem from a psychiatric disorder, vigorous treatment of the condition should be tried to eliminate the cognitive complaint or to confirm that it persists despite adequate resolution of the mood or anxiety symptoms. Patients with degenerative diseases may also be depressed or anxious, and those aspects of their condition often respond to therapy while not necessarily improving cognition. Antidepressants, such as selective serotonin reuptake inhibitors (SSRIs) or serotonin-norepinephrine reuptake inhibitors (SNRIs) **(Chap. 452)**, which feature anxiolytic properties but few cognitive side effects, provide the mainstay of treatment when necessary. Anticonvulsants are used to control AD-associated seizures.

Agitation, hallucinations, delusions, and confusion are difficult to treat. These behavioral problems represent major causes for nursing home placement and institutionalization. Before treating these behaviors with medications, the clinician should aggressively seek out modifiable environmental or metabolic factors. Hunger, lack of exercise, toothache, constipation, urinary tract or respiratory infection, electrolyte imbalance, and drug toxicity all represent easily correctable causes that can be remedied without psychoactive drugs. Drugs such as phenothiazines and benzodiazepines may ameliorate the behavior problems but have untoward side effects such as sedation, rigidity, or dyskinesia; benzodiazepines can occasionally produce paradoxical disinhibition. Despite their unfavorable side effect profile, second-generation antipsychotics such as quetiapine (starting dose, 12.5–25 mg daily) can be used for patients with agitation, aggression, and psychosis, although the risk profile for these compounds is significant, including increased mortality in patients with dementia. When patients do not respond to treatment, it is usually a mistake to advance to higher doses or to use anticholinergic drugs (like diphenhydramine) or sedatives (such as barbiturates or benzodiazepines). It is important to recognize and treat depression; treatment can begin with a low dose of an SSRI (e.g., escitalopram, starting dose 5 mg daily, target dose 5–10 mg daily) while monitoring for efficacy and toxicity. Sometimes apathy, visual hallucinations, depression, and other psychiatric symptoms respond to cholinesterase inhibitors, especially in DLB, obviating the need for other more toxic therapies.

Cholinesterase inhibitors are being used to treat AD (donepezil, rivastigmine, galantamine) and PDD (rivastigmine). Memantine is useful for some patients with moderate to severe AD; its major benefit relates to decreasing caregiver burden, most likely by decreasing resistance to dressing and grooming support. In moderate to severe AD, the combination of memantine and a cholinesterase inhibitor delayed nursing home placement in several studies, although other studies have not supported the efficacy of adding memantine to the regimen. Memantine should be used with great caution, or not at all, in patients with DLB, due to risk of worsening agitation and confusion. Therapies targeting the production, aggregation, and spread of misfolded proteins associated with dementia are under development. Recently the first drug in this class, the amyloid-beta targeting monoclonal antibody aducanumab, was approved by the United States Food & Drug Administration for treatment of Alzheimer's disease (Chap. 431). Other drugs under development target disease-associated neuroinflammation metabolic changes, synaptic loss, and neurotransmitter changes.

Proactive approaches reduce the occurrence of delirium in hospitalized patients. Frequent orientation, cognitive activities, sleep-enhancement measures, vision and hearing aids, and correction of dehydration are all valuable in decreasing the likelihood of delirium.

Nondrug behavior therapy has an important place in dementia management. The primary goals are to make the patient's life comfortable, uncomplicated, and safe. Preparing lists, schedules, calendars, and labels can be helpful in the early stages. It is also useful to stress familiar routines, walks, and simple physical exercises. For many demented patients, memory for events is worse than their ability to carry out routine activities, and they may still be able to take part in their favorite hobbies, sports, and social activities. Demented patients often object to losing control over familiar tasks such as driving, cooking, and handling finances. Attempts to help may be greeted with complaints, depression, or anger. Hostile responses on the part of the caregiver are counterproductive and sometimes even harmful. Reassurance, distraction, and calm positive statements are more productive when resistance is present. Eventually, tasks such as finances and driving must be assumed by others, and the patient will conform and adjust. Safety is an important issue that includes not only driving but controlling the kitchen, bathroom, and sleeping area environments, as well as stairways. These areas need to be monitored, supervised, and made as safe as possible. A move to a retirement complex, assisted-living center, or nursing home can initially increase confusion and agitation. Repeated reassurance, reorientation, and careful introduction to the new personnel will help to smooth the process. Providing activities that are known to be enjoyable to the patient can also help.

The clinician must pay special attention to frustration and depression among family members and caregivers. Caregiver guilt and burnout are common. Family members often feel overwhelmed and helpless and may vent their frustrations on the patient, each other, and health care providers. Caregivers should be encouraged to take advantage of day-care facilities and respite services. Education and counseling about dementia

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FURTHER READING

Access Provided by:
SILVERCHAIR

common. Family members often feel overwhelmed and helpless and may vent their frustrations on the patient, each other, and health care providers. Caregivers should be encouraged to take advantage of day-care facilities and respite services. Education and counseling about dementia are important. Local and national support groups, such as the Alzheimer's Association (*www.alz.org*), can provide considerable help.

FURTHER READING

Barton C et al: Non-pharmacological management of behavioral symptoms in frontotemporal and other dementias. Curr Neurol Neurosci Rep 16:14, 2016. [PubMed: 26750129]

Griem J et al: Psychologic/functional forms of memory disorder. Handb Clin Neurol 139:407, 2017.

Wesley SF, Ferguson D: Autoimmune encephalitides and rapidly progressive dementias. Semin Neurol 39:283, 2019. [PubMed: 30925620]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 29: Dementia, William W. Seeley; Gil D. Rabinovici; Bruce L. Miller
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders

M.-Marsel Mesulam

INTRODUCTION

The cerebral cortex of the human brain contains ~20 billion neurons spread over an area of 2.5 m^2. The primary sensory and motor areas constitute 10% of the cerebral cortex. The rest is subsumed by modality-selective, heteromodal, paralimbic, and limbic areas collectively known as the *association cortex* **(Fig. 30-1)**. The association cortex mediates the integrative processes that subserve cognition, emotion, and comportment. A systematic testing of these mental functions is necessary for the effective clinical assessment of the association cortex and its diseases. According to current thinking, there are no centers for "hearing words," "perceiving space," or "storing memories." Cognitive and behavioral functions (domains) are coordinated by intersecting *large-scale neural networks* that contain interconnected cortical and subcortical components. Five anatomically defined *large-scale networks* are most relevant to clinical practice: (1) a left-dominant perisylvian network for language, (2) a right-dominant parietofrontal network for spatial orientation, (3) an occipitotemporal network for face and object recognition, (4) a limbic network for episodic memory and emotional modulation, and (5) a prefrontal network for the executive control of cognition and comportment. Investigations based on functional imaging have also identified a *default mode network*, which becomes activated when the person is not engaged in a specific task requiring attention to external events. The clinical consequences of damage to this network are not yet fully defined.

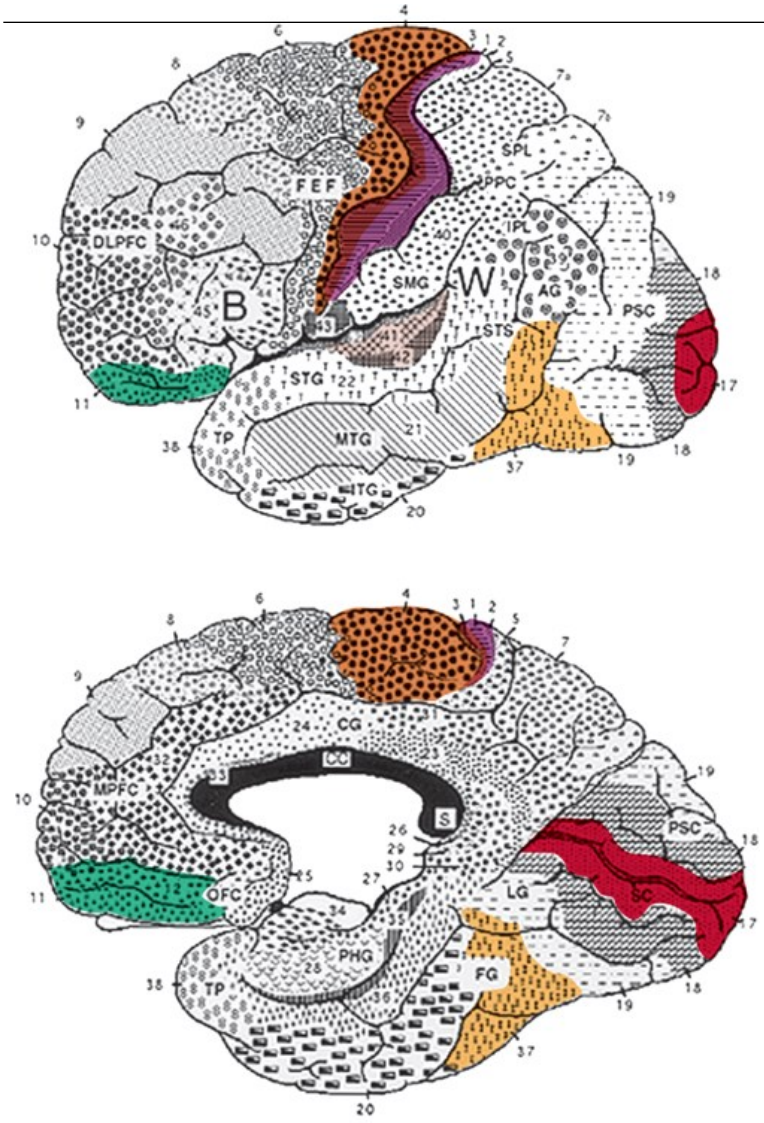
FIGURE 30-1

Lateral (*top*) and medial (*bottom*) views of the cerebral hemispheres. The numbers refer to the Brodmann cytoarchitectonic designations. Area 17 corresponds to the primary visual cortex, 41–42 to the primary auditory cortex, 1–3 to the primary somatosensory cortex, and 4 to the primary motor cortex. The rest of the cerebral cortex contains association areas. AG, angular gyrus; B, Broca's area; CC, corpus callosum; CG, cingulate gyrus; DLPFC, dorsolateral prefrontal cortex; FEF, frontal eye fields (premotor cortex); FG, fusiform gyrus; IPL, inferior parietal lobule; ITG, inferior temporal gyrus; LG, lingual gyrus; MPFC, medial prefrontal cortex; MTG, middle temporal gyrus; OFC, orbitofrontal cortex; PHG, parahippocampal gyrus; PPC, posterior parietal cortex; PSC, peristriate cortex; SC, striate cortex; SMG, supramarginal gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus; STS, superior temporal sulcus; TP, temporopolar cortex; W, Wernicke's area.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

THE LEFT PERISYLVIAN NETWORK FOR LANGUAGE AND APHASIAS

The production and comprehension of words and sentences is dependent on the integrity of a distributed network located along the perisylvian region of the language-dominant (usually left) hemisphere. One hub, situated in the inferior frontal gyrus, is known as *Broca's area*. Damage to this region impairs fluency of verbal output and the grammatical structure of sentences. The location of a second hub, critical for language comprehension, is less clearly settled. Accounts of patients with focal cerebrovascular lesions identified *Wernicke's area*, located at the parietotemporal junction, as a critical hub for word and sentence comprehension. Occlusive or embolic strokes involving this area interfere with the ability to understand spoken or written language as well as the ability to express thoughts through meaningful words and statements. However, investigations of patients with the neurodegenerative syndrome of primary progressive aphasia (PPA) have shown that sentence comprehension is a widely distributed faculty jointly subserved by Broca's and Wernicke's areas, and that the areas critical for word comprehension are more closely associated with the anterior temporal lobe than with Wernicke's area. All components of the language network are interconnected with each other and with surrounding parts of the frontal, parietal, and temporal lobes. Damage to this network gives rise to language impairments known as aphasia. Aphasia should be diagnosed only when there are deficits in the formal aspects of language, such as word finding, word choice, comprehension, spelling, or grammar. Dysarthria, apraxia of speech, and mutism do not by themselves lead to a diagnosis of aphasia. In ~90% of right-handers and 60% of left-handers, aphasia occurs only after lesions of the left hemisphere.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

A deficit of naming (*anomia*) is the single most common finding in aphasic patients. When asked to name a common object, the patient may fail to come up with the appropriate word, may provide a circumlocutious description of the object ("the thing for writing"), or may come up with the wrong

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

CLINICAL EXAMINATION

The clinical examination of language should include the assessment of naming, spontaneous speech, comprehension, repetition, reading, and writing. A deficit of naming (*anomia*) is the single most common finding in aphasic patients. When asked to name a common object, the patient may fail to come up with the appropriate word, may provide a circumlocutious description of the object ("the thing for writing"), or may come up with the wrong word (*paraphasia*). If the patient offers an incorrect but related word ("pen" for "pencil"), the naming error is known as a *semantic paraphasia*; if the word approximates the correct answer but is phonetically inaccurate ("plentil" for "pencil"), it is known as a *phonemic paraphasia*. In most anomias, the patient cannot retrieve the appropriate name when shown an object but can point to the appropriate object when the name is provided by the examiner. This is known as a one-way (or retrieval-based) naming deficit. A two-way (comprehension-based or semantic) naming deficit exists if the patient can neither provide nor recognize the correct name. *Spontaneous speech* is described as "fluent" if it maintains appropriate output volume, phrase length, and melody or as "nonfluent" if it is sparse and halting and average utterance length is below four words. The examiner also should note the integrity of *grammar* as manifested by word order (syntax), tenses, suffixes, prefixes, plurals, and possessives. *Comprehension* can be tested by assessing the patient's ability to follow conversation, asking yes-no questions ("Can a dog fly?" "Does it snow in summer?"), asking the patient to point to appropriate objects ("Where is the source of illumination in this room?"), or asking for verbal definitions of single words. *Repetition* is assessed by asking the patient to repeat single words, short sentences, or strings of words such as "No ifs, ands, or buts." The testing of repetition with tongue twisters such as "hippopotamus" and "Irish constabulary" provides a better assessment of dysarthria and apraxia of speech than of aphasia. It is important to make sure that the number of words does not exceed the patient's attention span. Otherwise, the failure of repetition becomes a reflection of the narrowed attention span (auditory working memory) rather than an indication of an aphasic deficit caused by dysfunction of a hypothetical *phonological loop* in the language network. *Reading* should be assessed for deficits in reading aloud as well as comprehension. *Alexia* describes an inability to either read aloud or comprehend written words and sentences; *agraphia* (or dysgraphia) is used to describe an acquired deficit in spelling.

Aphasias can arise acutely in cerebrovascular accidents (CVAs) or gradually in neurodegenerative diseases. In CVAs, damage encompasses cerebral cortex as well as deep white matter pathways interconnecting otherwise unaffected cortical areas. The syndromes listed in **Table 30-1** are most applicable to this group, where gray matter and white matter at the lesion site are abruptly and jointly destroyed. Progressive neurodegenerative diseases can have cellular, laminar, and regional specificity for the cerebral cortex, giving rise to a different set of aphasias that will be described separately.

TABLE 30-1

Clinical Features of Aphasias and Related Conditions Commonly Seen in Cerebrovascular Accidents

	COMPREHENSION	REPETITION OF SPOKEN LANGUAGE	NAMING	FLUENCY
Wernicke's	Impaired	Impaired	Impaired	Preserved or increased
Broca's	Preserved (except grammar)	Impaired	Impaired	Decreased
Global	Impaired	Impaired	Impaired	Decreased
Conduction	Preserved	Impaired	Impaired	Preserved
Nonfluent (anterior) transcortical	Preserved	Preserved	Impaired	Impaired
Fluent (posterior) transcortical	Impaired	Preserved	Impaired	Preserved
Isolation	Impaired	Echolalia	Impaired	No purposeful speech
Anomic	Preserved	Preserved	Impaired	Preserved except for word-finding
	language			

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 30-1

Clinical Features of Aphasias and Related Conditions Commonly Seen in Cerebrovascular Accidents

	COMPREHENSION	REPETITION OF SPOKEN LANGUAGE	NAMING	FLUENCY
Wernicke's	Impaired	Impaired	Impaired	Preserved or increased
Broca's	Preserved (except grammar)	Impaired	Impaired	Decreased
Global	Impaired	Impaired	Impaired	Decreased
Conduction	Preserved	Impaired	Impaired	Preserved
Nonfluent (anterior) transcortical	Preserved	Preserved	Impaired	Impaired
Fluent (posterior) transcortical	Impaired	Preserved	Impaired	Preserved
Isolation	Impaired	Echolalia	Impaired	No purposeful speech
Anomic	Preserved	Preserved	Impaired	Preserved except for word-finding pauses
Pure word deafness	Impaired only for spoken language	Impaired	Preserved	Preserved
Pure alexia	Impaired only for reading	Preserved	Preserved	Preserved

Wernicke's Aphasia

Comprehension is impaired for spoken and written words and sentences. Language output is fluent but is highly paraphasic and circumlocutious. Paraphasic errors may lead to strings of neologisms, which lead to "jargon aphasia." Speech contains few substantive nouns. The output is therefore voluminous but uninformative. For example, a patient attempts to describe how his wife accidentally threw away something important, perhaps his dentures: "We don't need it anymore, she says. And with it when that was downstairs was my teeth-tick … a … den … dentith … my dentist. And they happened to be in that bag … see? …Where my two … two little pieces of dentist that I use … that I … all gone. If she throws the whole thing away … visit some friends of hers and she can't throw them away."

Gestures and pantomime do not improve communication. The patient may not realize that his or her language is incomprehensible and may appear angry and impatient when the examiner fails to decipher the meaning of a severely paraphasic statement. In some patients, this type of aphasia can be associated with severe agitation and paranoia. The ability to follow commands aimed at axial musculature may be preserved. The dissociation between the failure to understand simple questions ("What is your name?") in a patient who rapidly closes his or her eyes, sits up, or rolls over when asked to do so is characteristic of Wernicke's aphasia and helps differentiate it from deafness, psychiatric disease, or malingering. Patients with Wernicke's aphasia cannot express their thoughts in meaning-appropriate words and cannot decode the meaning of words in any modality of input. This aphasia therefore has expressive as well as receptive components. Repetition, naming, reading, and writing also are impaired.

The lesion site most commonly associated with Wernicke's aphasia caused by CVAs is the posterior portion of the language network. An embolus to the inferior division of the middle cerebral artery (MCA), to the posterior temporal or angular branches in particular, is the most common etiology (Chap. 426). Intracerebral hemorrhage, head trauma, and neoplasm are other causes of Wernicke's aphasia. A coexisting right hemianopia or superior

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

usually settle the issue. Prognosis for recovery of language function is guarded.

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

inferior division of the middle cerebral artery (MCA), to the posterior temporal or angular branches in particular, is the most common etiology (Chap. 426). Intracerebral hemorrhage, head trauma, and neoplasm are other causes of Wernicke's aphasia. A coexisting right hemianopia or superior quadrantanopia is common, and mild right nasolabial flattening may be found, but otherwise, the examination is often unrevealing. The paraphasic, neologistic speech in an agitated patient with an otherwise unremarkable neurologic examination may lead to the suspicion of a primary psychiatric disorder such as schizophrenia or mania, but the other components characteristic of acquired aphasia and the absence of prior psychiatric disease usually settle the issue. Prognosis for recovery of language function is guarded.

Broca's Aphasia

Speech is nonfluent, labored, interrupted by many word-finding pauses, and usually dysarthric. It is impoverished in function words but enriched in meaning-appropriate nouns. Abnormal word order and the inappropriate deployment of *bound morphemes* (word endings used to denote tenses, possessives, or plurals) lead to a characteristic agrammatism. Speech is telegraphic and pithy but quite informative. In the following passage, a patient with Broca's aphasia describes his medical history: "I see … the dotor, dotor sent me … Bosson. Go to hospital. Dotor … kept me beside. Two, tee days, doctor send me home."

Output may be reduced to a grunt or single word ("yes" or "no"), which is emitted with different intonations in an attempt to express approval or disapproval. In addition to fluency, naming and repetition are impaired. Comprehension of spoken language is intact except for syntactically difficult sentences with a passive voice structure or embedded clauses, indicating that Broca's aphasia is not just an "expressive" or "motor" disorder and that it also may involve a comprehension deficit in decoding syntax. Patients with Broca's aphasia can be tearful, easily frustrated, and profoundly depressed. Insight into their condition is preserved, in contrast to Wernicke's aphasia. Even when spontaneous speech is severely dysarthric, the patient may be able to display a relatively normal articulation of words when singing. This dissociation has been used to develop specific therapeutic approaches (melodic intonation therapy) for Broca's aphasia. Additional neurologic deficits include right facial weakness, hemiparesis or hemiplegia, and a buccofacial apraxia characterized by an inability to carry out motor commands involving oropharyngeal and facial musculature (e.g., patients are unable to demonstrate how to blow out a match or suck through a straw). The cause is most often infarction of Broca's area (the inferior frontal convolution; "B" in Fig. 30-1) and surrounding anterior perisylvian and insular cortex due to occlusion of the superior division of the MCA (Chap. 426). Mass lesions, including tumor, intracerebral hemorrhage, and abscess, also may be responsible. When the cause of Broca's aphasia is stroke, recovery of language function generally peaks within 2–6 months, after which time further progress is limited. Speech therapy is more successful than in Wernicke's aphasia.

Conduction Aphasia

Speech output is fluent but contains many phonemic paraphasias, comprehension of spoken language is intact, and repetition is severely impaired. Naming elicits phonemic paraphasias, and spelling is impaired. Reading aloud is impaired, but reading comprehension is preserved. The responsible lesion, usually a CVA in the temporoparietal or dorsal perisylvian region, interferes with the function of the phonological loop interconnecting Broca's area with Wernicke's area. Occasionally, a transient Wernicke's aphasia may rapidly resolve into a conduction aphasia. The paraphasic and circumlocutious output in conduction aphasia interferes with the ability to express meaning, but this deficit is not nearly as severe as the one displayed by patients with Wernicke's aphasia. Associated neurologic signs in conduction aphasia vary according to the primary lesion site.

Transcortical Aphasias: Fluent and Nonfluent

Clinical features of *fluent (posterior) transcortical aphasia* are similar to those of Wernicke's aphasia, but repetition is intact. The lesion site disconnects the intact core of the language network from other temporoparietal association areas. Associated neurologic findings may include hemianopia. Cerebrovascular lesions (e.g., infarctions in the posterior watershed zone) and neoplasms that involve the temporoparietal cortex posterior to Wernicke's area are common causes. The features of *nonfluent (anterior) transcortical aphasia* are similar to those of Broca's aphasia, but repetition is intact and agrammatism is less pronounced. The neurologic examination may be otherwise intact, but a right hemiparesis also can exist. The lesion site disconnects the intact language network from prefrontal areas of the brain and usually involves the anterior watershed zone between anterior and MCA territories or the supplementary motor cortex in the territory of the anterior cerebral artery.

Global and Isolation Aphasias

Global aphasia represents the combined dysfunction of Broca's and Wernicke's areas and usually results from strokes that involve the entire MCA distribution in the left hemisphere. Speech output is nonfluent, and comprehension of language is severely impaired. Related signs include right hemiplegia, hemisensory loss, and homonymous hemianopia. *Isolation aphasia* represents a combination of the two transcortical aphasias.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

frontal, parietal, and temporal cortex. Lesions are patchy and can be associated with anoxia, carbon monoxide poisoning, or complete watershed zone

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

distribution in the left hemisphere. Speech output is nonfluent, and comprehension of language is severely impaired. Related signs include right hemiplegia, hemisensory loss, and homonymous hemianopia. *Isolation aphasia* represents a combination of the two transcortical aphasias. Comprehension is severely impaired, and there is no purposeful speech output. The patient may parrot fragments of heard conversations (*echolalia*), indicating that the neural mechanisms for repetition are at least partially intact. This condition represents the pathologic function of the language network when it is isolated from other regions of the brain. Broca's and Wernicke's areas tend to be spared, but there is damage to the surrounding frontal, parietal, and temporal cortex. Lesions are patchy and can be associated with anoxia, carbon monoxide poisoning, or complete watershed zone infarctions.

Anomic Aphasia

This form of aphasia may be considered the "minimal dysfunction" syndrome of the language network. Articulation, comprehension, and repetition are intact, but confrontation naming, word finding, and spelling are impaired. Word-finding pauses are uncommon, so language output is fluent but paraphasic, circumlocutious, and uninformative. The lesion sites can be anywhere within the left hemisphere language network, including the middle and inferior temporal gyri. *Anomic aphasia is the single most common language disturbance seen in head trauma, metabolic encephalopathy, and Alzheimer's disease.*

Pure Word Deafness

The most common causes are either bilateral or left-sided MCA strokes affecting the superior temporal gyrus. The net effect of the underlying lesion is to interrupt the flow of information from the auditory association cortex to the language network. Patients have no difficulty understanding written language and can express themselves well in spoken or written language. They have no difficulty interpreting and reacting to environmental sounds if the primary auditory cortex and auditory association areas of the right hemisphere are spared. Because auditory information cannot be conveyed to the language network, however, it cannot be decoded into neural word representations, and the patient reacts to speech as if it were in an alien tongue that cannot be deciphered. Patients cannot repeat spoken language but have no difficulty naming objects. In time, patients with pure word deafness teach themselves lipreading and may appear to have improved. There may be no additional neurologic findings, but agitated paranoid reactions are common in the acute stages. Cerebrovascular lesions are the most common cause.

Pure Alexia Without Agraphia

This is the visual equivalent of pure word deafness. The lesions (usually a combination of damage to the left occipital cortex and to a posterior sector of the corpus callosum—the splenium) interrupt the flow of visual input into the language network. There is usually a right hemianopia, but the core language network remains unaffected. The patient can understand and produce spoken language, name objects in the left visual hemifield, repeat, and write. However, the patient acts as if illiterate when asked to read even the simplest sentence because the visual information from the written words (presented to the intact left visual hemifield) cannot reach the language network. Objects in the left hemifield may be named accurately because they activate nonvisual associations in the right hemisphere, which in turn can access the language network through transcallosal pathways anterior to the splenium. Patients with this syndrome also may lose the ability to name colors, although they can match colors. This is known as a *color anomia.* The most common etiology of pure alexia is a vascular lesion in the territory of the posterior cerebral artery or an infiltrating neoplasm in the left occipital cortex that involves the optic radiations as well as the crossing fibers of the splenium. Because the posterior cerebral artery also supplies medial temporal components of the limbic system, a patient with pure alexia also may experience an amnesia, but this is usually transient because the limbic lesion is unilateral.

Apraxia and Aphemia

Apraxia designates a complex motor deficit that cannot be attributed to pyramidal, extrapyramidal, cerebellar, or sensory dysfunction and that does not arise from the patient's failure to understand the nature of the task. *Apraxia of speech* is used to designate articulatory abnormalities in the duration, fluidity, and stress of syllables that make up words. It can arise with CVAs in the posterior part of Broca's area or in the course of frontotemporal lobar degeneration (FTLD) with tauopathy. *Aphemia* is a severe form of acute speech apraxia that presents with severely impaired fluency (often mutism). Recovery is the rule and involves an intermediate stage of hoarse whispering. Writing, reading, and comprehension are intact, and so this is not a true aphasic syndrome. CVAs in parts of Broca's area or subcortical lesions that undercut its connections with other parts of the brain may be present. Occasionally, the lesion site is on the medial aspects of the frontal lobes and may involve the supplementary motor cortex of the left hemisphere. *Ideomotor apraxia* is diagnosed when commands to perform a specific motor act ("cough," "blow out a match") or pantomime the use of a common tool (a comb, hammer, straw, or toothbrush) in the absence of the real object cannot be followed. The patient's ability to

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

represent a disconnection of the language network from pyramidal motor systems so that commands to execute complex movements are understood but cannot be conveyed to the appropriate motor areas. *Buccofacial apraxia* involves apraxic deficits in movements of the face and mouth. Ideomotor

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

fluency (often mutism). Recovery is the rule and involves an intermediate stage of hoarse whispering. Writing, reading, and comprehension are intact, use of a common tool (a comb, hammer, straw, or toothbrush) in the absence of the real object cannot be followed. The patient's ability to comprehend the command is ascertained by demonstrating multiple movements and establishing that the correct one can be recognized. Some patients with this type of apraxia can imitate the appropriate movement when it is demonstrated by the examiner and show no impairment when handed the real object, indicating that the sensorimotor mechanisms necessary for the movement are intact. Some forms of ideomotor apraxia represent a disconnection of the language network from pyramidal motor systems so that commands to execute complex movements are understood but cannot be conveyed to the appropriate motor areas. *Buccofacial apraxia* involves apraxic deficits in movements of the face and mouth. Ideomotor *limb apraxia* encompasses apraxic deficits in movements of the arms and legs. Ideomotor apraxia almost always is caused by lesions in the left hemisphere and is commonly associated with aphasic syndromes, especially Broca's aphasia and conduction aphasia. Because the handling of real objects is not impaired, ideomotor apraxia by itself causes no major limitation of daily living activities. Patients with lesions of the anterior corpus callosum can display ideomotor apraxia confined to the left side of the body, a sign known as *sympathetic dyspraxia.* A severe form of sympathetic dyspraxia, known as the *alien hand* syndrome, is characterized by additional features of motor disinhibition on the left hand. *Ideational apraxia* refers to a deficit in the sequencing of goal-directed movements in patients who have no difficulty executing the individual components of the sequence. For example, when the patient is asked to pick up a pen and write, the sequence of uncapping the pen, placing the cap at the opposite end, turning the point toward the writing surface, and writing may be disrupted, and the patient may be seen trying to write with the wrong end of the pen or even with the removed cap. These motor sequencing problems usually are seen in the context of confusional states and dementias rather than focal lesions associated with aphasic conditions. *Limb-kinetic apraxia* involves clumsiness in the use of tools or objects that cannot be attributed to sensory, pyramidal, extrapyramidal, or cerebellar dysfunction. This condition can emerge in the context of focal premotor cortex lesions or *corticobasal degeneration* and can interfere with the use of tools and utensils.

Gerstmann's Syndrome

The combination of *acalculia* (impairment of simple arithmetic), *dysgraphia* (impaired writing), *finger anomia* (an inability to name individual fingers such as the index and thumb), and *right-left confusion* (an inability to tell whether a hand, foot, or arm of the patient or examiner is on the right or left side of the body) is known as Gerstmann's syndrome. In making this diagnosis, it is important to establish that the finger and left-right naming deficits are not part of a more generalized anomia and that the patient is not otherwise aphasic. When Gerstmann's syndrome arises acutely and in isolation, it is commonly associated with damage to the inferior parietal lobule (especially the angular gyrus) in the left hemisphere.

Pragmatics and Prosody

Pragmatics refers to aspects of language that communicate attitude, affect, and the figurative rather than literal aspects of a message (e.g., "green thumb" does not refer to the actual color of the finger). One component of pragmatics, *prosody,* refers to variations of melodic stress and intonation that influence attitude and the inferential aspect of verbal messages. For example, the two statements "He *is* clever." and "He is *clever*?" contain an identical word choice and syntax but convey vastly different messages because of differences in the intonation with which the statements are uttered. Damage to right hemisphere regions corresponding to Broca's area impairs the ability to introduce meaning-appropriate prosody into spoken language. The patient produces grammatically correct language with accurate word choice, but the statements are uttered in a monotone that interferes with the ability to convey the intended stress and effect. Patients with this type of *aprosodia* give the mistaken impression of being depressed or indifferent. Other aspects of pragmatics, especially the ability to infer the figurative aspect of a message, become impaired by damage to the right hemisphere or frontal lobes.

Subcortical Aphasia

Damage to subcortical components of the language network (e.g., the striatum and thalamus of the left hemisphere) also can lead to aphasia. The resulting syndromes contain combinations of deficits in the various aspects of language but rarely fit the specific patterns described in Table 30-1. In a patient with a CVA, an anomic aphasia accompanied by dysarthria or a fluent aphasia with hemiparesis should raise the suspicion of a subcortical lesion site.

CLINICAL PRESENTATION AND DIAGNOSIS OF PPA

Aphasias caused by CVAs start suddenly and display maximal deficits at the onset. These are the "classic" aphasias described above. Aphasias caused by neurodegenerative diseases have an insidious onset and relentless progression. The neuropathology can be selective not only for gray matter but also for specific layers and cell types. The clinico-anatomic patterns are therefore different from those described in Table 30-1.

Several neurodegenerative syndromes, such as typical Alzheimer-type (amnestic; **Chap. 431**) and frontotemporal (behavioral; **Chap. 432**) dementias, Alzheimer's disease (AD) pathology. Rarely, an identical syndrome can be caused by Creutzfeldt-Jacob disease (CJD) but with a more rapid progression (**Chap. 438**)

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

Several neurodegenerative syndromes, such as typical Alzheimer-type (amnestic; **Chap. 431**) and frontotemporal (behavioral; **Chap. 432**) dementias, can also include language impairments as the disease progresses. In these cases, the aphasia is an ancillary component of the overall syndrome. A diagnosis of primary progressive aphasia (PPA) is justified only if the language disorder (i.e., aphasia) arises in relative isolation, becomes the primary concern that brings the patient to medical attention, and remains the most salient deficit for 1–2 years. PPA can be caused by either FTLD or Alzheimer's disease (AD) pathology. Rarely, an identical syndrome can be caused by Creutzfeldt-Jacob disease (CJD) but with a more rapid progression **(Chap. 438)**.

LANGUAGE IN PPA

The impairments of language in PPA have slightly different patterns from those seen in CVA-caused aphasias. For example, the full syndrome of Wernicke's aphasia is almost never seen in PPA, confirming the view that sentence comprehension and word comprehension are controlled by different regions of the language network. Three major subtypes of PPA can be recognized.

Agrammatic PPA

The *agrammatic variant* is characterized by consistently low fluency and impaired grammar but intact word comprehension. It most closely resembles Broca's aphasia or anterior transcortical aphasia but usually lacks the right hemiparesis or dysarthria and may have more profound impairments of grammar. Peak sites of neuronal loss (gray matter atrophy) include the left inferior frontal gyrus where Broca's area is located. The neuropathology is usually a FTLD with tauopathy but can also be an atypical form of AD pathology.

Semantic PPA

The *semantic variant* is characterized by preserved fluency and syntax but poor single-word comprehension and profound two-way naming impairments. This kind of aphasia is not seen with CVAs. It differs from Wernicke's aphasia or posterior transcortical aphasia because speech is usually informative and repetition is intact. Comprehension of sentences is relatively preserved if the meaning is not too dependent on words that fail to be understood allowing the patient to surmise the gist of the conversation through contextual cues. Such patients may appear unimpaired in the course of casual small talk but become puzzled upon encountering an undecipherable word such as "pumpkin" or "umbrella." Peak atrophy sites are located in the left anterior temporal lobe, indicating that this part of the brain plays a critical role in the comprehension of words, especially words that denote concrete objects. This is a part of the brain that was not included within the classic language network, probably because it is not a common site for focal CVAs. The neuropathology is frequently an FTLD with abnormal precipitates of the 43-kDa transactive response DNA-binding protein TDP-43 of type C.

Logopenic PPA

The *logopenic variant* is characterized by preserved syntax and comprehension but frequent and severe word-finding pauses, anomia, circumlocutions, and simplifications during spontaneous speech. Repetition is usually impaired. Peak atrophy sites are located in the temporoparietal junction and posterior temporal lobe, partially overlapping with traditional location of Wernicke's area. However, the comprehension impairment of *Wernicke's aphasia* is absent probably because the underlying deep white matter, frequently damaged by CVAs, remains relatively intact in PPA. The repetition impairment suggests that parts of Wernicke's area are critical for phonological loop functionality. In contrast to Broca's aphasia or agrammatic PPA, the interruption of fluency is variable so that speech may appear entirely normal if the patient is allowed to engage in small talk. Logopenic PPA resembles the anomic aphasia of Table 30-1 but usually has longer and more frequent word-finding pauses. When repetition is impaired, the aphasia resembles the *conduction aphasia* in Table 30-1. Of all PPA subtypes, this is the one most commonly associated with the pathology of AD, but FTLD can also be the cause. In addition to these three major subtypes, there is also a *mixed* type of PPA where grammar, fluency, and word comprehension are jointly impaired. This is most like the global aphasia of Table 30-1. Rarely, PPA can present with patterns reminiscent of *pure word deafness* or *Gerstmann's syndrome*.

THE PARIETOFRONTAL NETWORK FOR SPATIAL ORIENTATION

Adaptive spatial orientation is subserved by a large-scale network containing three major cortical components. The *cingulate cortex* provides access to a motivational mapping of the extrapersonal space, the *posterior parietal cortex* to a sensorimotor representation of salient extrapersonal events, and the *frontal eye fields* to motor strategies for attentional behaviors **(Fig. 30-2)**. Subcortical components of this network include the striatum and the thalamus. Damage to this network can undermine the distribution of attention within the extrapersonal space, giving rise to hemispatial neglect

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FIGURE 30-2

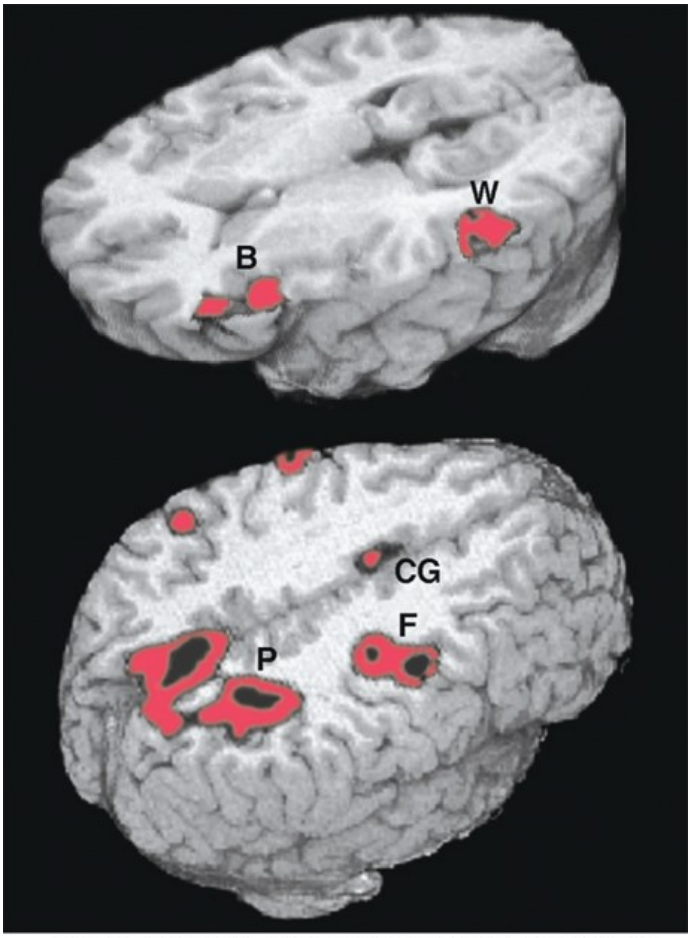
ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

the *frontal eye fields* to motor strategies for attentional behaviors **(Fig. 30-2)**. Subcortical components of this network include the striatum and the thalamus. Damage to this network can undermine the distribution of attention within the extrapersonal space, giving rise to hemispatial neglect, simultanagnosia, and object finding failures. The integration of egocentric (self-centered) with allocentric (object-centered) coordinates can also be disrupted, giving rise to impairments in route finding, the ability to avoid obstacles, and the ability to dress.

FIGURE 30-2

Functional magnetic resonance imaging of language and spatial attention in neurologically intact subjects. The red and black areas show regions of task-related significant activation. (*Top*) The subjects were asked to determine if two words were synonymous. This language task led to the simultaneous activation of the two components of the language network, Broca's area (B) and Wernicke's area (W). The activations are exclusively in the left hemisphere. (*Bottom*) The subjects were asked to shift spatial attention to a peripheral target. This task led to the simultaneous activation of the three epicenters of the attentional network: the posterior parietal cortex (P), the frontal eye fields (F), and the cingulate gyrus (CG). The activations are predominantly in the right hemisphere. *(Courtesy of Darren Gitelman, MD.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

HEMISPATIAL NEGLECT

Contralesional hemispatial neglect represents one outcome of damage to the cortical or subcortical components of this network. *The traditional view that hemispatial neglect always denotes a parietal lobe lesion is inaccurate.* According to one model of spatial cognition, the right hemisphere directs attention within the *entire* extrapersonal space, whereas the left hemisphere directs attention mostly within the contralateral right hemispace. Consequently, left hemisphere lesions do not give rise to much contralesional neglect because the global attentional mechanisms of the right hemisphere can compensate for the loss of the *contralaterally* directed attentional functions of the left hemisphere. Right hemisphere lesions, however, give rise to severe contralesional left hemispatial neglect because the unaffected left hemisphere does not contain ipsilateral attentional mechanisms. This model is consistent with clinical experience, which shows that contralesional neglect is more common, more severe, and longer

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

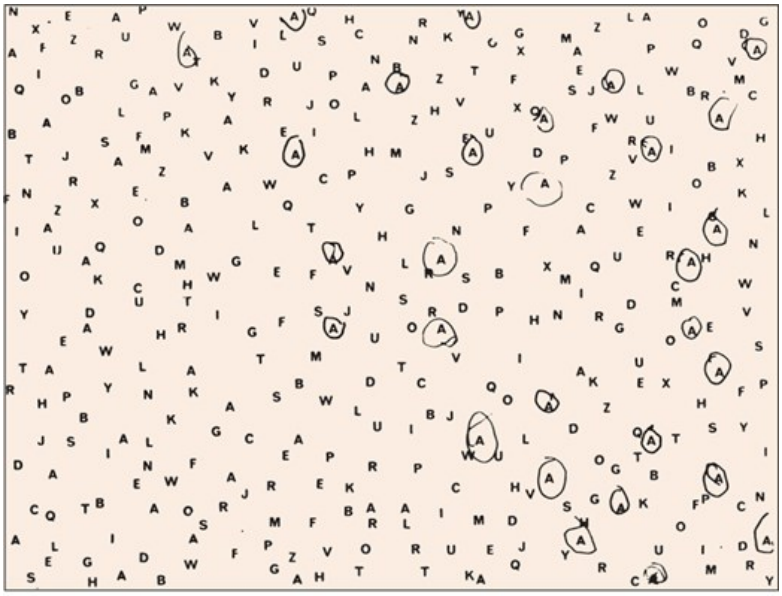
that hemispatial neglect always denotes a parietal lobe lesion is inaccurate. According to one model of spatial cognition, the right hemisphere directs however, give rise to severe contralesional left hemispatial neglect because the unaffected left hemisphere does not contain ipsilateral attentional mechanisms. This model is consistent with clinical experience, which shows that contralesional neglect is more common, more severe, and longer lasting after damage to the right hemisphere than after damage to the left hemisphere. Severe neglect for the right hemispace is rare, even in left-handers with left hemisphere lesions.

Clinical Examination

Patients with severe neglect may fail to dress, shave, or groom the left side of the body; fail to eat food placed on the left side of the tray; and fail to read the left half of sentences. When asked to copy a simple line drawing, the patient fails to copy detail on the left, and when the patient is asked to write, there is a tendency to leave an unusually wide margin on the left. Two bedside tests that are useful in assessing neglect are *simultaneous bilateral stimulation* and *visual target cancellation*. In the former, the examiner provides either unilateral or simultaneous bilateral stimulation in the visual, auditory, and tactile modalities. After right hemisphere injury, patients who have no difficulty detecting unilateral stimuli on either side experience the bilaterally presented stimulus as coming only from the right. This phenomenon is known as *extinction* and is a manifestation of the sensory-representational aspect of hemispatial neglect. In the target detection task, targets (e.g., A's) are interspersed with foils (e.g., other letters of the alphabet) on a 21.5- to 28.0-cm (8.5–11 in.) sheet of paper, and the patient is asked to circle all the targets. A failure to detect targets on the left is a manifestation of the exploratory (motor) deficit in hemispatial neglect **(Fig. 30-3A)**. Hemianopia is not by itself sufficient to cause the target detection failure because the patient is free to turn the head and eyes to the left. Target detection failures therefore reflect a distortion of spatial attention, not just of sensory input. Some patients with neglect also may deny the existence of hemiparesis and may even deny ownership of the paralyzed limb, a condition known as *anosognosia*.

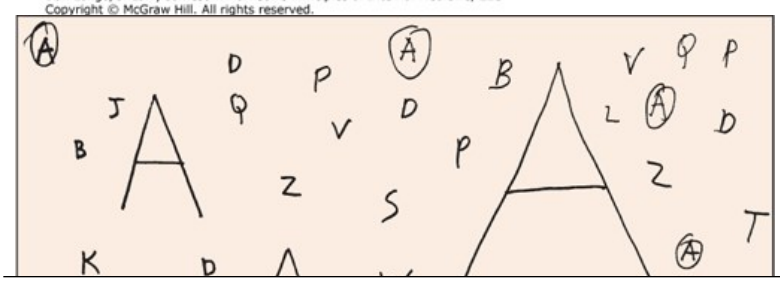
FIGURE 30-3

A. A 47-year-old man with a large frontoparietal lesion in the right hemisphere was asked to circle all the A's. Only targets on the right are circled. This is a manifestation of left hemispatial neglect. ***B.*** A 70-year-old woman with a 2-year history of degenerative dementia was able to circle most of the small targets but ignored the larger ones. This is a manifestation of simultanagnosia.



A

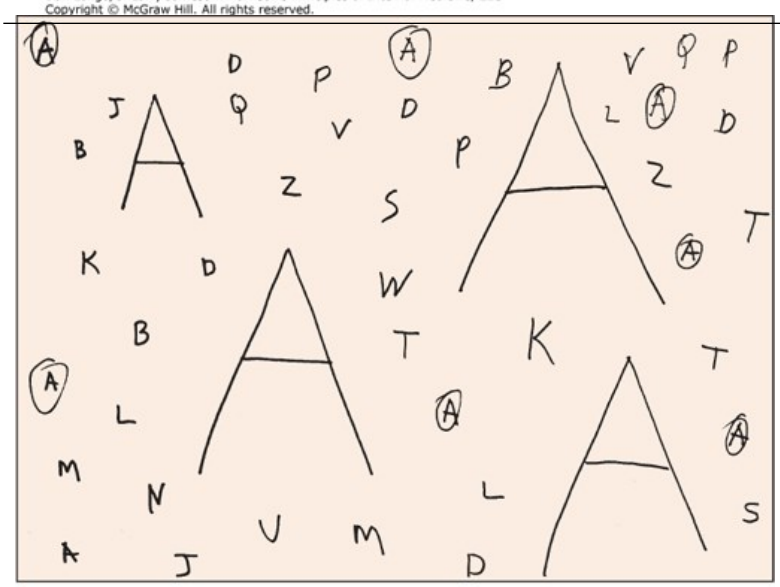
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.



B

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser,
Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

BÁLINT'S SYNDROME, SIMULTANAGNOSIA, DRESSING APRAXIA, CONSTRUCTION APRAXIA, AND ROUTE-FINDING IMPAIRMENTS

Bilateral involvement of the network for spatial attention, especially its parietal components, leads to a state of severe spatial disorientation known as *Bálint's syndrome*. Bálint's syndrome involves deficits in the orderly visuomotor scanning of the environment (*oculomotor apraxia*), accurate manual reaching toward visual targets (*optic ataxia*), and the ability to integrate visual information in the center of gaze with more peripheral information (*simultanagnosia*). A patient with simultanagnosia "misses the forest for the trees." For example, a patient who is shown a table lamp and asked to name the object may look at its circular base and call it an ashtray. Some patients with simultanagnosia report that objects they look at may vanish suddenly, probably indicating an inability to compute the oculomotor return to the original point of gaze after brief saccadic displacements. Movement and distracting stimuli greatly exacerbate the difficulties of visual perception. Simultanagnosia can occur without the other two components of Bálint's syndrome, especially in association with AD.

A modification of the letter cancellation task described above can be used for the bedside diagnosis of simultanagnosia. In this modification, some of the targets (e.g., A's) are made to be much larger than the others (7.5–10 cm vs 2.5 cm [3–4 in. vs 1 in.] in height), and all targets are embedded among foils. Patients with simultanagnosia display a counterintuitive but characteristic tendency to miss the larger targets **(Fig. 30-3B)**. This occurs because the information needed for the identification of the larger targets cannot be confined to the immediate line of gaze and requires the integration of visual information across multiple fixation points. The greater difficulty in the detection of the larger targets also indicates that poor acuity is not responsible for the impairment of visual function and that the problem is central rather than peripheral. The test shown in Fig. 30-3B is not by itself sufficient to diagnose simultanagnosia as some patients with a frontal network syndrome may omit the letters that appear incongruous for the size of the paper. This may happen because they lack the mental flexibility to realize that the two types of targets are symbolically identical despite being superficially different.

Bilateral parietal lesions can impair the integration of egocentric with allocentric spatial coordinates. One manifestation is *dressing apraxia*. A patient with this condition is unable to align the body axis with the axis of the garment and can be seen struggling as he or she holds a coat from its bottom or extends his or her arm into a fold of the garment rather than into its sleeve. Lesions that involve the posterior parietal cortex also lead to severe difficulties in copying simple line drawings. This is known as a *construction apraxia* and is much more severe if the lesion is in the right hemisphere. In some patients with right hemisphere lesions, the drawing difficulties are confined to the left side of the figure and represent a manifestation of hemispatial neglect; in others, there is a more universal deficit in reproducing contours and three-dimensional perspective. Impairments of route finding can be included in this group of disorders, which reflect an inability to orient the self with respect to external objects and landmarks.

Causes of Spatial Disorientation and the Posterior Cortical Atrophy Syndrome

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

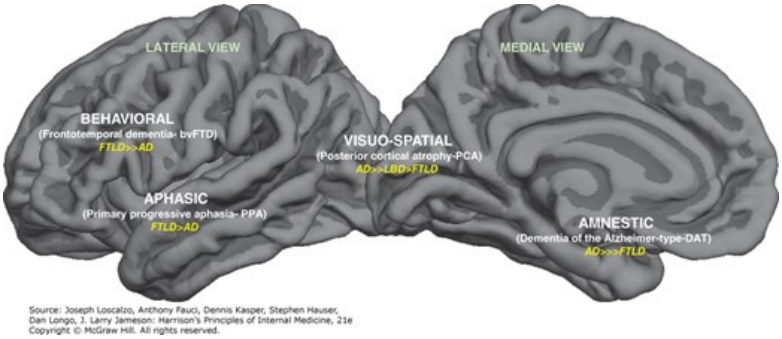
these patients display considerable improvement of hemispatial neglect, usually within the first several weeks. Bálint's syndrome, dressing apraxia, and route-finding impairments are more likely to result from bilateral dorsal parietal lesions; common settings for acute onset include watershed

Cerebrovascular lesions and neoplasms in the right hemisphere are common causes of hemispatial neglect. Depending on the site of the lesion, a patient with neglect also may have hemiparesis, hemihypesthesia, and hemianopia on the left, but these are not invariant findings. The majority of these patients display considerable improvement of hemispatial neglect, usually within the first several weeks. Bálint's syndrome, dressing apraxia, and route-finding impairments are more likely to result from bilateral dorsal parietal lesions; common settings for acute onset include watershed infarction between the middle and posterior cerebral artery territories, hypoglycemia, and sagittal sinus thrombosis.

A progressive form of spatial disorientation, known as the *posterior cortical atrophy* (PCA) syndrome, most commonly represents a variant of AD with unusual concentrations of neurofibrillary degeneration in the parieto-occipital cortex and the superior colliculus **(Fig. 30-4)**. Lewy body disease (LBD), CJD, and FTLD (corticobasal degeneration type) are other possible causes. The patient displays progressive hemispatial neglect, Bálint's syndrome, and route-finding impairments, usually accompanied by dressing and construction apraxia.

FIGURE 30-4

Four focal dementia syndromes and their most likely neuropathologic correlates. AD, Alzheimer's disease; bvFTD, behavioral variant frontotemporal dementia; DAT, amnestic dementia of the Alzheimer type; FTLD, frontotemporal lobar degeneration (tau or TDP-43 type); LBD, Lewy body disease; PCA, posterior cortical atrophy syndrome; PPA, primary progressive aphasia.



THE OCCIPITOTEMPORAL NETWORK FOR FACE AND OBJECT RECOGNITION

A patient with *prosopagnosia* cannot recognize familiar faces, including, sometimes, the reflection of their own face in the mirror. This is not a perceptual deficit because prosopagnosic patients easily can tell whether two faces are identical. Furthermore, a prosopagnosic patient who cannot recognize a familiar face by visual inspection alone can use auditory cues to reach appropriate recognition if allowed to listen to the person's voice. The deficit in prosopagnosia is therefore modality-specific and reflects the existence of a lesion that prevents the activation of otherwise intact multimodal associative templates by relevant visual input. Prosopagnosic patients characteristically have no difficulty with the generic identification of a face as a face or a car as a car, but may not recognize the identity of an individual face or the make of an individual car. This reflects a visual recognition deficit for proprietary features that characterize individual members of an object class. When recognition problems become more generalized and extend to the generic identification of common objects, the condition is known as *visual object agnosia*. A patient with anomia cannot name the object but can describe its use. In contrast, a patient with visual agnosia is unable either to name a visually presented object or to describe its use. Face and object recognition disorders also can result from the simultanagnosia of Bálint's syndrome, in which case they are known as *apperceptive* agnosias as opposed to the *associative* agnosias that result from inferior temporal lobe lesions.

CAUSES AND RELATION TO SEMANTIC DEMENTIA

The characteristic lesions in prosopagnosia and visual object agnosia of acute onset consist of bilateral infarctions in the territory of the posterior cerebral arteries that involve the fusiform gyrus. Associated deficits can include visual field defects (especially superior quadrantanopias) and a centrally based color blindness known as achromatopsia. Rarely, the responsible lesion is unilateral. In such cases, prosopagnosia is associated with lesions in the right hemisphere, and object agnosia with lesions in the left. Degenerative diseases of anterior and inferior temporal cortex can cause progressive associative prosopagnosia and object agnosia. The combination of progressive associative agnosia and a fluent aphasia with word comprehension impairment is known as *semantic dementia*. Patients with semantic dementia fail to recognize faces and objects and cannot understand the meaning of words denoting objects. This needs to be differentiated from the semantic type of PPA where there is severe impairment in

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

LIMBIC NETWORK FOR EXPLICIT MEMORY AND AMNESIA

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

comprehension impairment is known as *semantic dementia*. Patients with semantic dementia fail to recognize faces and objects and cannot understand the meaning of words denoting objects. This needs to be differentiated from the semantic type of PPA where there is severe impairment in understanding words that denote objects and in naming faces and objects but a relative preservation of face and object recognition. The anterior temporal lobe atrophy is usually bilateral in semantic dementia whereas it tends to affect mostly the left hemisphere in semantic PPA. Acute onset of the semantic dementia syndrome can be associated with herpes simplex encephalitis.

LIMBIC NETWORK FOR EXPLICIT MEMORY AND AMNESIA

Limbic areas (e.g., the hippocampus, amygdala, and entorhinal cortex), paralimbic areas (e.g., the cingulate gyrus, insula, temporopolar cortex, and parts of orbitofrontal regions), the anterior and medial nuclei of the thalamus, the medial and basal parts of the striatum, and the hypothalamus collectively constitute a distributed network known as the *limbic system*. The behavioral affiliations of this network can be classified into two groups. One includes the coordination of emotion, motivation, affiliative behaviors, autonomic tone, and endocrine function. These functions are under the influence of the amygdala and anterior paralimbic areas. They make up the salience network. The two neurologic conditions that most frequently interfere with this group of limbic functions are temporal lobe epilepsy and behavioral variant frontotemporal dementia (bvFTD). An additional area of specialization for the limbic network and the one that is of most relevance to clinical practice is that of declarative (explicit) memory for recent episodes and experiences. This function is under the influence of the hippocampus, entorhinal cortex, posterior paralimbic areas, and limbic nuclei of the thalamus. This part of the limbic system is also known as the Papez circuit. A disturbance of explicit memory is known as an *amnestic state*. In the absence of deficits in motivation, attention, language, or visuospatial function, the clinical diagnosis of a persistent global amnestic state is always associated with bilateral damage to the limbic network, usually within the hippocampo-entorhinal complex or the thalamus. Damage to the limbic network does not necessarily destroy memories but interferes with their conscious recall in coherent form. The individual fragments of information remain preserved despite the limbic lesions and can sustain what is known as *implicit memory*. For example, patients with amnestic states can acquire new motor or perceptual skills even though they may have no conscious knowledge of the experiences that led to the acquisition of these skills.

The memory disturbance in the amnestic state is multimodal and includes retrograde and anterograde components. The *retrograde amnesia* involves an inability to recall experiences that occurred before the onset of the amnestic state. Relatively recent events are more vulnerable to retrograde amnesia than are more remote and more extensively consolidated events. A patient who comes to the emergency room complaining that he cannot remember his or her identity but can remember the events of the previous day almost certainly does not have a neurologic cause of memory disturbance. The second and most important component of the amnestic state is the *anterograde amnesia*, which indicates an inability to store, retain, and recall new knowledge. Patients with amnestic states cannot remember what they ate a few hours ago or the details of an important event they may have experienced in the recent past. In the acute stages, there also may be a tendency to fill in memory gaps with inaccurate, fabricated, and often implausible information. This is known as *confabulation*. Patients with the amnestic syndrome forget that they forget and tend to deny the existence of a memory problem when questioned. Confabulation is more common in cases where the underlying lesion also interferes with parts of the frontal network, as in the case of the Wernicke-Korsakoff syndrome or traumatic head injury.

CLINICAL EXAMINATION

A patient with an amnestic state is almost always disoriented, especially to time, and has little knowledge of current news. The anterograde component of an amnestic state can be tested with a list of four to five words read aloud by the examiner up to five times or until the patient can immediately repeat the entire list without an intervening delay. The next phase of the recall occurs after a period of 5–10 min during which the patient is engaged in other tasks. Amnestic patients fail this phase of the task and may even forget that they were given a list of words to remember. Accurate recognition of the words by multiple choice in a patient who cannot recall them indicates a less severe memory disturbance that affects mostly the retrieval stage of memory. The retrograde component of an amnesia can be assessed with questions related to autobiographical or historic events. The anterograde component of amnestic states is usually much more prominent than the retrograde component. In rare instances, occasionally associated with temporal lobe epilepsy or herpes simplex encephalitis, the retrograde component may dominate. Confusional states caused by toxic-metabolic encephalopathies and some types of frontal lobe damage lead to secondary memory impairments, especially at the stages of encoding and retrieval, even in the absence of limbic lesions. This sort of memory impairment can be differentiated from the amnestic state by the presence of additional impairments in the attention-related tasks described below in the section on the frontal lobes.

CAUSES, INCLUDING ALZHEIMER'S DISEASE

Neurologic diseases that give rise to an amnestic state include tumors (of the sphenoid wing, posterior corpus callosum, thalamus, or medial temporal lobe), infarctions (in the territories of the anterior or posterior cerebral arteries), head trauma, herpes simplex encephalitis, Wernicke-Korsakoff

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

underlying neurologic disease, the patient also may have visual field deficits, eye movement limitations, or cerebellar findings.

Neurologic diseases that give rise to an amnestic state include tumors (of the sphenoid wing, posterior corpus callosum, thalamus, or medial temporal lobe), infarctions (in the territories of the anterior or posterior cerebral arteries), head trauma, herpes simplex encephalitis, Wernicke-Korsakoff encephalopathy, autoimmune limbic encephalitis, and degenerative dementias such as AD and Pick's disease. The one common denominator of all these diseases is the presence of bilateral lesions within one or more components in the limbic network. Occasionally, unilateral left-sided hippocampal lesions can give rise to an amnestic state, but the memory disorder tends to be transient. Depending on the nature and distribution of the underlying neurologic disease, the patient also may have visual field deficits, eye movement limitations, or cerebellar findings.

The most common cause of progressive memory impairments in the elderly is AD. This is why a predominantly amnestic dementia is also known as a dementia of the Alzheimer type (DAT). A prodromal stage of DAT, when daily living activities are generally preserved, is known as amnestic mild cognitive impairment (MCI). The predilection of the entorhinal cortex and hippocampus for early neurofibrillary degeneration by typical AD pathology is responsible for the initially selective impairment of episodic memory. In time, additional impairments in language, attention, and visuospatial skills emerge as the neurofibrillary degeneration spreads to additional neocortical areas. Less frequently, amnestic dementias can also be caused by FTLD.

Transient global amnesia is a distinctive syndrome usually seen in late middle age. Patients become acutely disoriented and repeatedly ask who they are, where they are, and what they are doing. The spell is characterized by anterograde amnesia (inability to retain new information) and a retrograde amnesia for relatively recent events that occurred before the onset. The syndrome usually resolves within 24–48 h and is followed by the filling in of the period affected by the retrograde amnesia, although there is persistent loss of memory for the events that occurred during the ictus. Recurrences are noted in ~20% of patients. Migraine, temporal lobe seizures, and perfusion abnormalities in the posterior cerebral territory have been postulated as causes of transient global amnesia. The absence of associated neurologic findings occasionally may lead to the incorrect diagnosis of a psychiatric disorder.

THE PREFRONTAL NETWORK FOR EXECUTIVE FUNCTION AND BEHAVIOR

The frontal lobes can be subdivided into motor-premotor, dorsolateral prefrontal, medial prefrontal, and orbitofrontal components. The terms *frontal lobe syndrome* and *prefrontal cortex* refer only to the last three of these four components. These are the parts of the cerebral cortex that show the greatest phylogenetic expansion in primates, especially in humans. The dorsolateral prefrontal, medial prefrontal, and orbitofrontal areas, along with the subcortical structures with which they are interconnected (i.e., the head of the caudate and the dorsomedial nucleus of the thalamus), collectively make up a large-scale network that coordinates exceedingly complex aspects of human cognition and behavior. The prefrontal network overlaps with the salience network through the anterior cingulate gyrus and parts of the orbitofrontal region. Impairments of social conduct and empathy seen in neurodegenerative frontal dementias (such as bvFTD) are attributed to pathology of the prefrontal and salience networks.

The prefrontal network plays an important role in behaviors that require multitasking and the integration of thought with emotion. Cognitive operations impaired by prefrontal cortex lesions often are referred to as "executive functions." The most common clinical manifestations of damage to the prefrontal network take the form of two relatively distinct syndromes. In the *frontal abulic syndrome*, the patient shows a loss of initiative, creativity, and curiosity and displays a pervasive emotional blandness, apathy, and lack of empathy. In the *frontal disinhibition syndrome*, the patient becomes socially disinhibited and shows severe impairments of judgment, insight, foresight, and the ability to mind rules of conduct. The dissociation between intact intellectual function and a total lack of even rudimentary common sense is striking. Despite the preservation of all essential memory functions, the patient cannot learn from experience and continues to display inappropriate behaviors without appearing to feel emotional pain, guilt, or regret when those behaviors repeatedly lead to disastrous consequences. The impairments may emerge only in real-life situations when behavior is under minimal external control and may not be apparent within the structured environment of the medical office. Testing judgment by asking patients what they would do if they detected a fire in a theater or found a stamped and addressed envelope on the road is not very informative because patients who answer these questions wisely in the office may still act very foolishly in real-life settings. The physician must therefore be prepared to make a diagnosis of frontal lobe disease based on historic information alone even when the mental state is quite intact in the office examination.

CLINICAL EXAMINATION

The emergence of developmentally primitive reflexes, also known as frontal release signs, such as grasping (elicited by stroking the palm) and sucking (elicited by stroking the lips) are seen primarily in patients with large structural lesions that extend into the premotor components of the frontal lobes or in the context of metabolic encephalopathies. The vast majority of patients with prefrontal lesions and frontal lobe behavioral syndromes do not display these reflexes. Damage to the frontal lobe disrupts a variety of attention-related functions, including working memory (the transient online holding and manipulation of information), concentration span, the effortful scanning and retrieval of stored information, the inhibition of immediate but inappropriate responses, and mental flexibility. Digit span (which should be seven forward and five reverse) is decreased, reflecting poor working

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

instruction is to raise the finger upon hearing one tap but keep it still upon hearing two taps), the patient shows a characteristic inability to inhibit the response to the "no go" stimulus. Mental flexibility (tested by the ability to shift from one criterion to another in sorting or matching tasks) is

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

but inappropriate responses, and mental flexibility. Digit span (which should be seven forward and five reverse) is decreased, reflecting poor working memory; the recitation of the months of the year in reverse order (which should take <15 s) is slowed as another indication of poor working memory; and the fluency in producing words starting with the letter a, f, or s that can be generated in 1 min (normally ≥12 per letter) is diminished even in nonaphasic patients, indicating an impairment in the ability to search and retrieve information from long-term stores. In "go–no go" tasks (where the instruction is to raise the finger upon hearing one tap but keep it still upon hearing two taps), the patient shows a characteristic inability to inhibit the response to the "no go" stimulus. Mental flexibility (tested by the ability to shift from one criterion to another in sorting or matching tasks) is impoverished; distractibility by irrelevant stimuli is increased; and there is a pronounced tendency for impersistence and perseveration. The ability for abstracting similarities and interpreting proverbs is also undermined.

The attentional deficits disrupt the orderly registration and retrieval of new information and lead to *secondary* deficits of explicit memory. The distinction of the underlying neural mechanisms is illustrated by the observation that severely amnestic patients who cannot remember events that occurred a few minutes ago may have intact if not superior working memory capacity as shown in tests of digit span. The use of the term *memory* to designate two completely different mental faculties is confusing. Working memory depends on the on-line holding of information for brief periods of time, whereas explicit memory depends on the off-line storage and subsequent retrieval of the information.

CAUSES: TRAUMA, NEOPLASM, AND FRONTOTEMPORAL DEMENTIA

The abulic syndrome tends to be associated with damage in dorsolateral or dorsomedial prefrontal cortex, and the disinhibition syndrome with damage in orbitofrontal or ventromedial cortex. These syndromes tend to arise almost exclusively after bilateral lesions. Unilateral lesions confined to the prefrontal cortex may remain silent until the pathology spreads to the other side; this explains why thromboembolic CVA is an unusual cause of the frontal lobe syndrome. When behavioral syndromes of the frontal network arise in conjunction with asymmetric disease, the lesion tends to be predominantly on the right side of the brain. Common settings for frontal lobe syndromes include head trauma, ruptured aneurysms, hydrocephalus, tumors (including metastases, glioblastoma, and falx or olfactory groove meningiomas), and focal degenerative diseases, especially FTLD. The most prominent neurodegenerative frontal syndrome is bvFTD. In many patients with bvFTD, the atrophy includes orbitofrontal cortex and also extends into the anterior temporal lobes, insula, and anterior cingulate cortex. Occasionally, atrophy predominantly in the right anterior temporal lobe presents with the bvFTD syndrome. The behavioral changes in these patients can range from apathy to shoplifting, compulsive gambling, sexual indiscretions, remarkable lack of common sense, new ritualistic behaviors, and alterations in dietary preferences, usually leading to increased taste for sweets or rigid attachment to specific food items. In many patients with AD, neurofibrillary degeneration eventually spreads to prefrontal cortex and gives rise to components of the frontal lobe syndrome, but almost always on a background of severe memory impairment. Rarely, the bvFTD syndrome can arise in isolation in the context of an atypical form of AD pathology.

Lesions in the caudate nucleus or in the dorsomedial nucleus of the thalamus (subcortical components of the prefrontal network) also can produce a frontal lobe syndrome affecting mostly executive functions. This is one reason why the changes in mental state associated with degenerative basal ganglia diseases such as Parkinson's disease and Huntington's disease display components of the frontal lobe syndrome. Bilateral multifocal lesions of the cerebral hemispheres, none of which are individually large enough to cause specific cognitive deficits such as aphasia and neglect, can collectively interfere with the connectivity and therefore integrating (executive) function of the prefrontal cortex. A frontal lobe syndrome, usually of the abulic form, is therefore the single most common behavioral profile associated with a variety of bilateral multifocal brain diseases, including metabolic encephalopathy, multiple sclerosis, and vitamin B_{12} deficiency, among others. Many patients with the clinical diagnosis of a frontal lobe syndrome tend to have lesions that do not involve prefrontal cortex but involve either the subcortical components of the prefrontal network or its connections with other parts of the brain. To avoid making a diagnosis of "frontal lobe syndrome" in a patient with no evidence of frontal cortex disease, it is advisable to use the diagnostic term *frontal network syndrome*, with the understanding that the responsible lesions can lie anywhere within this distributed network. A patient with frontal lobe disease raises potential dilemmas in differential diagnosis: the abulia and blandness may be misinterpreted as depression, and the disinhibition as idiopathic mania or acting out. Appropriate intervention may be delayed while a treatable tumor keeps expanding.

CARING FOR PATIENTS WITH DEFICITS OF HIGHER CEREBRAL FUNCTION

Spontaneous improvement of cognitive deficits following stroke or trauma is common. It is most rapid in the first few weeks but may continue for up to 2 years, especially in young individuals with single brain lesions. Some of the initial deficits in such cases appear to arise from remote dysfunction (diaschisis) in brain regions that are interconnected with the site of initial injury. Improvement in these patients may reflect, at least in part, a normalization of the remote dysfunction. Other mechanisms may involve functional reorganization in surviving neurons adjacent to the injury or the

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Pharmacologic and Nonpharmacologic Interventions

Some of the deficits described in this chapter are so complex that they may bewilder not only the patient and family but also the physician. The care of

normalization of the remote dysfunction. Other mechanisms may involve functional reorganization in surviving neurons adjacent to the injury or the compensatory use of homologous structures, e.g., the right superior temporal gyrus with recovery from Wernicke's aphasia. In contrast, neurodegenerative diseases show a progression of impairment but at rates that vary greatly from patient to patient.

Pharmacologic and Nonpharmacologic Interventions

Some of the deficits described in this chapter are so complex that they may bewilder not only the patient and family but also the physician. The care of patients with such deficits requires a careful evaluation of the history, cognitive test results, and diagnostic procedures. Each piece of information needs to be interpreted cautiously and placed in context. A complaint of "poor memory," for example, may reflect an anomia; poor scores on a learning task may reflect a weakness of attention rather than explicit memory; a report of depression or indifference may reflect impaired prosody rather than a change in mood or empathy; jocularity may arise from poor insight rather than good mood. Although there are few well-controlled studies, several nonpharmacologic interventions have been used to treat higher cortical deficits. These include speech therapy for aphasias, behavioral modification for compartmental disorders, and cognitive training for visuospatial disorientation and amnestic syndromes. More practical interventions, usually delivered through occupational therapy, aim to improve daily living activities through assistive devices and modifications of the home environment. Determining driving competence is challenging, especially in the early stages of dementing diseases. An on-the-road driving test and reports from family members may help time decisions related to this very important activity. In neurodegenerative conditions such as PPA, transcranial magnetic (or direct current) stimulation has had mixed success in eliciting symptomatic improvement. The goal is to activate remaining neurons at sites of atrophy or in unaffected regions of the contralateral hemisphere. Depression and sleep disorders can intensify the cognitive disorders and should be treated with appropriate modalities. If neuroleptics become absolutely necessary for the control of agitation, atypical neuroleptics are preferable because of their lower extrapyramidal side effects. Treatment with neuroleptics in elderly patients with dementia requires weighing the potential benefits against the potentially serious side effects. This is especially relevant to the case of patients with Lewy body dementia, who can be unusually sensitive to side effects.

As in all other branches of medicine, a crucial step in patient care is to identify the underlying cause of the impairment. This is easily done in cases of CVA, head trauma, or encephalitis but becomes particularly challenging in the dementias because the same progressive clinical syndrome can be caused by one of several neuropathologic entities. The advent of imaging, blood, and cerebrospinal fluid biomarkers now makes it possible to address this question with reasonable success and to make specific diagnoses of AD, LBD, CJD, and FTLD. A specific etiologic diagnosis allows the physician to recommend medications or clinical trials that are the most appropriate for the underlying disease process. A clinical assessment that identifies the principal domain of behavioral and cognitive impairment followed by the judicious use of biomarker information to surmise the nature of the underlying disease allows a personalized approach to patients with higher cognitive impairment.

FURTHER READING

Ghetti B et al: *Frontotemporal Dementias: Emerging Milestones of the 21st Century*. New York, Springer, 2021.

Henry ML et al: Retraining speech production and fluency in non-fluent/agrammatic primary progressive aphasia. Brain 141:1799, 2018. [PubMed: 29718131]

Mesulam M-M: Behavioral neuroanatomy: Large-scale networks, association cortex, frontal syndromes, the limbic system and hemispheric specialization, in *Principles of Behavioral and Cognitive Neurology*, (ed). New York, Oxford University Press, 2000, pp 1–120.

Mesulam M-M et al: Word comprehension in temporal cortex and Wernicke area: A PPA perspective. Neurology 92:e224, 2019. [PubMed: 30578374]

Miller BL, Boeve BF (eds): *The Behavioral Neurology of Dementia*, 2nd ed. Cambridge, Cambridge University Press, 2017.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 30: Aphasia, Memory Loss, and Other Cognitive Disorders, M.-Marsel Mesulam
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 31: Sleep Disorders

Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler

INTRODUCTION

Disturbed sleep is one of the most common health complaints that physicians encounter. More than one-half of adults in the United States experience at least intermittent sleep disturbance, and only 30% of adult Americans report consistently obtaining a sufficient amount of sleep. The National Academy of Medicine has estimated that 50–70 million Americans suffer from a chronic disorder of sleep and wakefulness, which can adversely affect daytime functioning as well as physical and mental health. A high prevalence of sleep disorders across all cultures is also now increasingly recognized, and these problems are expected to further increase in the years ahead as the global population ages. Over the last 30 years, the field of sleep medicine has emerged as a distinct specialty in response to the impact of sleep disorders and sleep deficiency on overall health. Nonetheless, over 80% of patients with sleep disorders remain undiagnosed and untreated—costing the U.S. economy over $400 billion annually in increased health care costs, lost productivity, accidents and injuries, and leading to the development of workplace-based sleep health education and sleep disorders screening programs designed to address this unmet medical need.

PHYSIOLOGY OF SLEEP AND WAKEFULNESS

Most adults need 7–9 h of sleep per night to promote optimal health, although the timing, duration, and internal structure of sleep vary among individuals. In the United States, adults tend to have one consolidated sleep episode each night, although in some cultures sleep may be divided into a mid-afternoon nap and a shortened night sleep. This pattern changes considerably over the life span, as infants and young children sleep considerably more than older people, while individuals >70 years of age sleep on average about an hour less than young adults.

The stages of human sleep are defined on the basis of characteristic patterns in the electroencephalogram (EEG), the electrooculogram (EOG—a measure of eye-movement activity), and the surface electromyogram (EMG) measured on the chin, neck, and legs. The continuous recording of these electrophysiologic parameters to define sleep and wakefulness is termed *polysomnography*.

Polysomnographic profiles define two basic states of sleep: (1) rapid eye movement (REM) sleep and (2) non–rapid eye movement (NREM) sleep. NREM sleep is further subdivided into three stages: N1, N2, and N3, characterized by an increasing threshold for arousal and slowing of the cortical EEG. REM sleep is characterized by a low-amplitude, mixed-frequency EEG, similar to NREM stage N1 sleep, and an EOG pattern of REMs that tend to occur in flurries or bursts. EMG activity is absent in nearly all skeletal muscles except those involved in respiration, reflecting the brainstem-mediated muscle paralysis that is characteristic of REM sleep.

ORGANIZATION OF HUMAN SLEEP

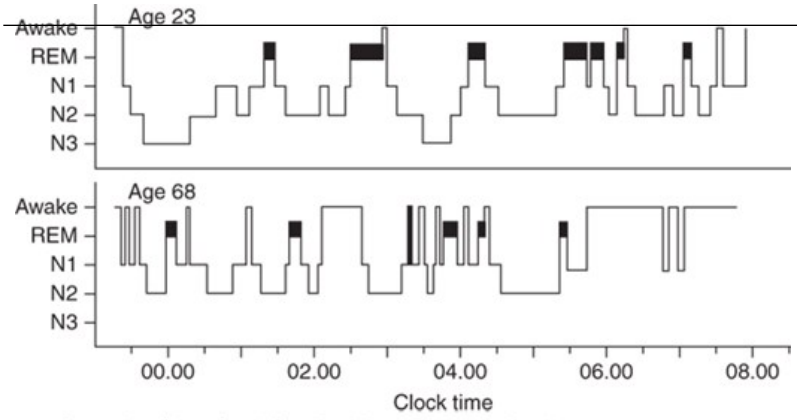
Normal nocturnal sleep in adults displays a consistent organization from night to night **(Fig. 31-1)**. After sleep onset, sleep usually progresses through NREM stages N1–N3 sleep within 45–60 min. NREM stage N3 sleep (also known as slow-wave sleep) predominates in the first third of the night and comprises 15–25% of total nocturnal sleep time in young adults. Sleep deprivation increases the rapidity of sleep onset and both the intensity and amount of slow-wave sleep.

FIGURE 31-1

Wake-sleep architecture. Alternating stages of wakefulness, the three stages of non–rapid eye movement sleep (N1–N3), and rapid eye movement (REM) sleep (*solid bars*) occur over the course of the night for representative young and older adult men. Characteristic features of sleep in older people include reduction of N3 slow-wave sleep, frequent spontaneous awakenings, early sleep onset, and early morning awakening.

Age 23

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The first REM sleep episode usually occurs in the second hour of sleep. NREM and REM sleep alternate through the night with an average period of 90–110 min (the "ultradian" sleep cycle). Overall, in a healthy young adult, REM sleep constitutes 20–25% of total sleep, and NREM stages N1 and N2 constitute 50–60%.

Age has a profound impact on sleep state organization **(Fig. 31-1)**. N3 sleep is most intense and prominent during childhood, decreasing with puberty and across the second and third decades of life. In older adults, N3 sleep may be completely absent, and the remaining NREM sleep typically becomes more fragmented, with frequent awakenings from NREM sleep. It is the increased frequency of awakenings, rather than a decreased ability to fall back asleep, that accounts for the increased wakefulness during the sleep episode in older people. While REM sleep may account for 50% of total sleep time in infancy, the percentage falls off sharply over the first postnatal year as a mature REM-NREM cycle develops; thereafter, REM sleep occupies about 25% of total sleep time.

Sleep deprivation degrades cognitive performance, particularly on tests that require continual vigilance. Paradoxically, older people are less vulnerable than young adults to the neurobehavioral performance impairment induced by acute sleep deprivation, maintaining their reaction time and sustaining vigilance with fewer lapses of attention. However, it is more difficult for older adults to obtain recovery sleep after staying awake all night, as the ability to sleep during the daytime declines with age.

After sleep deprivation, NREM sleep generally recovers first, followed by REM sleep. However, because REM sleep tends to be most prominent in the second half of the night, sleep truncation (e.g., by an alarm clock) results in selective REM sleep deprivation. This may increase REM sleep pressure to the point where the first REM sleep may occur much earlier in the nightly sleep episode. Because several disorders (see below) also cause sleep fragmentation, it is important that the patient have sufficient sleep opportunity (at least 8 h per night) for several nights prior to a diagnostic polysomnogram.

There is growing evidence that inadequate sleep in humans is associated with glucose intolerance that may contribute to the development of diabetes, obesity, and the metabolic syndrome, as well as impaired immune responses, accelerated atherosclerosis, and increased risk of cardiac disease, cognitive impairment, Alzheimer's disease, and stroke. For these reasons, the National Academy of Medicine declared sleep deficiency and sleep disorders "an unmet public health problem."

WAKE AND SLEEP ARE REGULATED BY BRAIN CIRCUITS

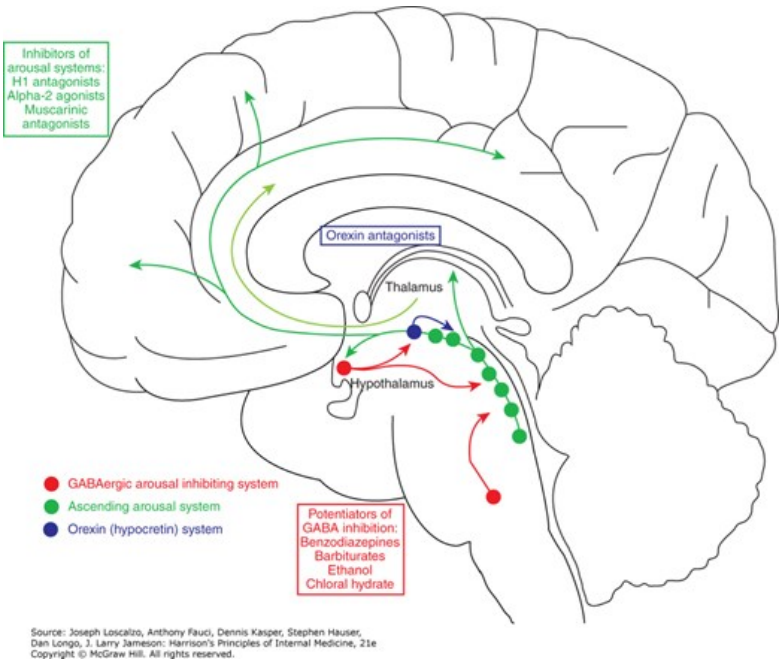
Two principal neural systems govern the expression of sleep and wakefulness. The ascending arousal system, illustrated in *green* in **Fig. 31-2**, consists of clusters of nerve cells extending from the upper pons to the hypothalamus and basal forebrain that activate the cerebral cortex, thalamus (which is necessary to relay sensory information to the cortex), and other forebrain regions. The ascending arousal neurons use monoamines (norepinephrine, dopamine, serotonin, and histamine), glutamate, or acetylcholine as neurotransmitters to activate their target neurons. Some basal forebrain neurons use γ-aminobutyric acid (GABA) to inhibit cortical inhibitory interneurons, thus promoting arousal. Additional wake-promoting neurons in the hypothalamus use the peptide neurotransmitter orexin (also known as hypocretin, shown in Fig. 31-2 in *blue*) to reinforce activity in the other arousal cell groups.

FIGURE 31-2

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

sleep-promoting system (*red*) consists of GABAergic neurons in the preoptic area and brainstem that inhibit the components of the arousal system, thus allowing sleep to occur. Drugs used to treat insomnia include those that block the effects of arousal system neurotransmitters (*green* and *blue*)

Relationship of drugs for insomnia with wake-sleep systems. The arousal system in the brain (*green*) includes monoaminergic, glutamatergic, and cholinergic neurons in the brainstem that activate neurons in the hypothalamus, thalamus, basal forebrain, and cerebral cortex. Orexin neurons (*blue*) in the hypothalamus, which are lost in narcolepsy, reinforce and stabilize arousal by activating other components of the arousal system. The sleep-promoting system (*red*) consists of GABAergic neurons in the preoptic area and brainstem that inhibit the components of the arousal system, thus allowing sleep to occur. Drugs used to treat insomnia include those that block the effects of arousal system neurotransmitters (*green* and *blue*) and those that enhance the effects of γ-aminobutyric acid (GABA) produced by the sleep system (*red*).



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Damage to the arousal system at the level of the rostral pons and lower midbrain causes coma, indicating that the ascending arousal influence from this level is critical in maintaining wakefulness. Injury to the hypothalamic branch of the arousal system causes profound sleepiness but usually not coma. Specific loss of the orexin neurons produces the sleep disorder narcolepsy (see below). Isolated damage to the thalamus causes loss of the content of wakefulness, known as a persistent vegetative state, but wake-sleep cycles are largely preserved.

The arousal system is turned off during sleep by inhibitory inputs from cell groups in the sleep-promoting system, shown in **Fig. 31-2** in *red*. These neurons in the preoptic area and pons use GABA to inhibit the arousal system. Additional neurons in the lateral hypothalamus containing the peptide melanin-concentrating hormone promote REM sleep. Many sleep-promoting neurons are themselves inhibited by inputs from the arousal system. This mutual inhibition between the arousal- and sleep-promoting systems forms a neural circuit akin to what electrical engineers call a “flip-flop switch.” A switch of this type tends to promote rapid transitions between the on (wake) and off (sleep) states, while avoiding intermediate states. The relatively rapid transitions between waking and sleeping states, as seen in the EEG of humans and animals, is consistent with this model.

Neurons in the ventrolateral preoptic nucleus, one of the key sleep-promoting sites, are lost during normal human aging, correlating with reduced ability to maintain sleep (sleep fragmentation). The ventrolateral preoptic neurons are also injured in Alzheimer’s disease, which may in part account for the poor sleep quality in those patients.

Transitions between NREM and REM sleep appear to be governed by a similar switch in the brainstem. GABAergic REM-Off neurons have been identified in the lower midbrain that inhibit REM-On neurons in the upper pons. The REM-On group contains both GABAergic neurons that inhibit the REM-Off group (thus satisfying the conditions for a REM sleep flip-flop switch) as well as glutamatergic neurons that project widely in the central nervous system (CNS) to cause the key phenomena associated with REM sleep. REM-On neurons that project to the medulla and spinal cord activate inhibitory (GABA and glycine-containing) interneurons, which in turn hyperpolarize the motor neurons, producing the paralysis of REM sleep. REM-On neurons that project to the forebrain may be important in producing dreams.

The REM sleep switch receives cholinergic input, which favors transitions to REM sleep, and monoaminergic (norepinephrine and serotonin) input that prevents REM sleep. As a result, drugs that increase monoamine tone (e.g., serotonin or norepinephrine reuptake inhibitors) tend to reduce the amount of REM sleep. Damage to the neurons that promote REM sleep paralysis can produce REM sleep behavior disorder, a condition in which

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The gradual increase in sleep drive with prolonged wakefulness, followed by deeper slow-wave sleep and prolonged sleep episodes, demonstrates

and glycine-containing) interneurons, which in turn hyperpolarize the motor neurons, producing the paralysis of REM sleep. REM-On neurons that prevents REM sleep. As a result, drugs that increase monoamine tone (e.g., serotonin or norepinephrine reuptake inhibitors) tend to reduce the amount of REM sleep. Damage to the neurons that promote REM sleep paralysis can produce REM sleep behavior disorder, a condition in which patients act out their dreams (see below).

SLEEP-WAKE CYCLES ARE DRIVEN BY HOMEOSTATIC, ALLOSTATIC, AND CIRCADIAN INPUTS

The gradual increase in sleep drive with prolonged wakefulness, followed by deeper slow-wave sleep and prolonged sleep episodes, demonstrates that there is a *homeostatic* mechanism that regulates sleep. The neurochemistry of sleep homeostasis is only partially understood, but with prolonged wakefulness, adenosine levels rise in parts of the brain. Adenosine may act through A1 receptors to directly inhibit many arousal-promoting brain regions. In addition, adenosine promotes sleep through A2a receptors; blockade of these receptors by caffeine is one of the chief ways in which people fight sleepiness. Other humoral factors, such as prostaglandin D_2, have also been implicated in this process. Both adenosine and prostaglandin D_2 activate the sleep-promoting neurons in the ventrolateral preoptic nucleus.

Allostasis is the physiologic response to a challenge such as physical danger or psychological threat that cannot be managed by homeostatic mechanisms. These stress responses can severely impact the need for and ability to sleep. For example, insomnia is very common in patients with anxiety and other psychiatric disorders. Stress-induced insomnia is even more common, affecting most people at some time in their lives. Positron emission tomography (PET) studies in patients with chronic insomnia show hyperactivation of components of the ascending arousal system, as well as their limbic system targets in the forebrain (e.g., cingulate cortex and amygdala). The limbic areas are not only targets for the arousal system, but they also send excitatory outputs back to the arousal system, which contributes to a vicious cycle of anxiety about insomnia that makes it more difficult to sleep. Approaches to treating insomnia may employ drugs that either inhibit the output of the ascending arousal system (*green* and *blue* in **Fig. 31-2**) or potentiate the output of the sleep-promoting system (*red* in **Fig. 31-2**). However, behavioral approaches (cognitive behavioral therapy [CBT] and sleep hygiene) that may reduce forebrain limbic activity at bedtime are often the best long-term treatment.

Sleep is also regulated by a strong *circadian* timing signal, driven by the suprachiasmatic nuclei (SCN) of the hypothalamus, as described below. The SCN sends outputs to key sites in the hypothalamus, which impose 24-h rhythms on a wide range of behaviors and body systems, including the wake-sleep cycle.

PHYSIOLOGY OF CIRCADIAN RHYTHMICITY

The wake-sleep cycle is the most evident of many 24-h rhythms in humans. Prominent daily variations also occur in endocrine, thermoregulatory, cardiac, pulmonary, renal, immune, gastrointestinal, and neurobehavioral functions. In evaluating daily rhythms in humans, it is important to distinguish between diurnal components passively evoked by periodic environmental or behavioral changes (e.g., the increase in blood pressure and heart rate that occurs upon assumption of the upright posture) and circadian rhythms actively driven by an endogenous oscillatory process (e.g., the circadian variations in adrenal cortisol and pineal melatonin secretion that persist across a variety of environmental and behavioral conditions).

At the cellular level, endogenous circadian rhythmicity is driven by self-sustaining feedback loops. While it is now recognized that most cells in the body have circadian clocks that regulate diverse physiologic processes, these clocks in different tissues, or even in different cells in the same tissue, when placed in isolation in a tissue explant are unable to maintain the long-term synchronization with each other that is required to produce useful 24-h rhythms aligned with the external light-dark cycle. The only tissue that maintains this rhythm in vitro is the SCN, whose neurons are interconnected with one another in such a way as to produce a near-24-h synchronous rhythm of neural activity even in prolonged slice culture. SCN neurons are located just above the optic chiasm in the hypothalamus, from which they receive visual input to synchronize them with the external world, and they have outputs to transmit that signal to the rest of the body. Bilateral destruction of the SCN results in a loss of most endogenous circadian rhythms including wake-sleep behavior and rhythms in endocrine and metabolic systems. The genetically determined period of this endogenous neural oscillator, which averages ~24.15 h in humans, is normally synchronized to the 24-h period of the environmental light-dark cycle through direct input from intrinsically photosensitive ganglion cells in the retina to the SCN. Humans are exquisitely sensitive to the resetting effects of light, particularly the shorter wavelengths (~460–500 nm) in the blue part of the visible spectrum. Small differences in circadian period contribute to variations in diurnal preference. Changes in homeostatic sleep regulation may underlie age-related changes in sleep-wake timing.

The timing and internal architecture of sleep are directly coupled to the output of the endogenous circadian pacemaker. Paradoxically, the endogenous circadian rhythm for wake propensity peaks just before the habitual bedtime, whereas that of sleep propensity peaks near the habitual wake time. These rhythms are thus timed to oppose the rise of sleep tendency throughout the usual waking day and the decline of sleep propensity during the habitual sleep episode, respectively, thus promoting consolidated sleep and wakefulness. Misalignment of the endogenous circadian pacemaker with the desired wake-sleep cycle can, therefore, induce insomnia, decrease alertness, and impair performance, posing health problems

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Polysomnographic staging of sleep correlates with behavioral changes during specific states and stages. During the transitional state (stage N1)

ACCESS Medicine®
Access Provided by: SILVERCHAIR

during the habitual sleep episode, respectively, thus promoting consolidated sleep and wakefulness. Misalignment of the endogenous circadian pacemaker with the desired wake-sleep cycle can, therefore, induce insomnia, decrease alertness, and impair performance, posing health problems for night-shift workers and airline travelers.

BEHAVIORAL AND PHYSIOLOGIC CORRELATES OF SLEEP STATES AND STAGES

Polysomnographic staging of sleep correlates with behavioral changes during specific states and stages. During the transitional state (stage N1) between wakefulness and deeper sleep, individuals may respond to faint auditory or visual signals. Formation of short-term memories is inhibited at the onset of NREM stage N1 sleep, which may explain why individuals aroused from that transitional sleep stage frequently lack situational awareness. After sleep deprivation, such transitions may intrude upon behavioral wakefulness notwithstanding attempts to remain continuously awake (for example, see "Shift-Work Disorder," below).

Subjects awakened from REM sleep recall vivid dream imagery >80% of the time, especially later in the night. Less vivid imagery may also be reported after NREM sleep interruptions. Certain disorders may occur during specific sleep stages and are described below under "Parasomnias." These include sleepwalking, night terrors, and enuresis (bed wetting), which occur most commonly in children during deep (N3) NREM sleep, and REM sleep behavior disorder, which occurs mainly among older men who fail to maintain full paralysis during REM sleep, and often call out, thrash around, or even act out fragments of dreams.

All major physiologic systems are influenced by sleep. Blood pressure and heart rate decrease during NREM sleep, particularly during N3 sleep. During REM sleep, bursts of eye movements are associated with large variations in both blood pressure and heart rate mediated by the autonomic nervous system. Cardiac dysrhythmias may occur selectively during REM sleep. Respiratory function also changes. In comparison to relaxed wakefulness, respiratory rate becomes slower but more regular during NREM sleep (especially N3 sleep) and becomes irregular during bursts of eye movements in REM sleep. Decreases in minute ventilation during NREM sleep are out of proportion to the decrease in metabolic rate, resulting in a slightly higher P_{CO_2}.

Within the brain itself, neurotransmission is supported by ion gradients across the cell membranes of neurons and astrocytes. These ion flows are accompanied by increases in intracellular volume, so that during wake, there is very little extracellular space in the brain. During sleep, intracellular volume is reduced, resulting in increased extracellular space, which has higher calcium and lower potassium concentrations, supporting hyperpolarization and reduced firing of neurons. This expansion of the extracellular space during sleep increases diffusion of substances that accumulate extracellularly, like β-amyloid peptide, enhancing their clearance from the brain via cerebrospinal fluid (CSF) flow. Recent evidence suggests that lack of adequate sleep may contribute to extracellular accumulation of β-amyloid peptide, a key step in the pathogenesis of Alzheimer's disease.

Endocrine function also varies with sleep. N3 sleep is associated with secretion of growth hormone in men, while sleep in general is associated with augmented secretion of prolactin in both men and women. Sleep has a complex effect on the secretion of luteinizing hormone (LH): during puberty, sleep is associated with increased LH secretion, whereas sleep in postpubertal women inhibits LH secretion in the early follicular phase of the menstrual cycle. Sleep onset (and probably N3 sleep) is associated with inhibition of thyroid-stimulating hormone and of the adrenocorticotropic hormone–cortisol axis, an effect that is superimposed on the prominent circadian rhythms in the two systems.

The pineal hormone melatonin is secreted predominantly at night in both day- and night-active species, reflecting the direct modulation of pineal activity by the SCN via the sympathetic nervous system, which innervates the pineal gland. Melatonin secretion does not require sleep, but melatonin secretion is inhibited by ambient light, an effect mediated by the neural connection from the retina to the pineal gland via the SCN. In humans, sleep efficiency is highest when sleep coincides with endogenous melatonin secretion. When endogenous melatonin levels are low, such as during the biological day or at the desired bedtime in people with delayed sleep-wake phase disorder (DSWPD), administration of exogenous melatonin can hasten sleep onset and increase sleep efficiency, but it does not increase sleep efficiency if administered when endogenous melatonin levels are elevated. This may explain why melatonin is often ineffective in the treatment of patients with primary insomnia. On the other hand, patients with sympathetic denervation of the pineal gland, such as occurs in cervical spinal cord injury or in patients with Parkinson's disease, often have low melatonin levels, and administration of melatonin (3 mg 30 min before bedtime) may help them sleep.

Sleep is accompanied by alterations of thermoregulatory function. NREM sleep is associated with an increase in the firing of warm-responsive neurons in the preoptic area and a fall in body temperature; conversely, skin warming without increasing core body temperature has been found to increase NREM sleep. REM sleep is associated with reduced thermoregulatory responsiveness.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

DISORDERS OF SLEEP AND WAKEFULNESS

APPROACH TO THE PATIENT

Sleep Disorders

Patients may seek help from a physician because of: (1) sleepiness or tiredness during the day; (2) difficulty initiating or maintaining sleep at night (insomnia); or (3) unusual behaviors during sleep itself (parasomnias).

Obtaining a careful history is essential. In particular, the duration, severity, and consistency of the symptoms are important, along with the patient's estimate of the consequences of the sleep disorder on waking function. Information from a bed partner or family member is often helpful because some patients may be unaware of symptoms such as heavy snoring or may underreport symptoms such as falling asleep at work or while driving. Physicians should inquire about when the patient typically goes to bed, when they fall asleep and wake up, whether they awaken during sleep, whether they feel rested in the morning, and whether they nap during the day. Depending on the primary complaint, it may be useful to ask about snoring, witnessed apneas, restless sensations in the legs, movements during sleep, depression, anxiety, and behaviors around the sleep episode. The physical examination may provide evidence of a small airway, large tonsils, or a neurologic or medical disorder that contributes to the main complaint.

It is important to remember that, rarely, seizures may occur exclusively during sleep, mimicking a primary sleep disorder; such sleep-related seizures typically occur during episodes of NREM sleep and may take the form of generalized tonic-clonic movements (sometimes with urinary incontinence or tongue biting) or stereotyped movements in partial complex epilepsy **(Chap. 418)**.

It is often helpful for the patient to complete a daily sleep log for 1–2 weeks to define the timing and amounts of sleep. When relevant, the log can also include information on levels of alertness, work times, and drug and alcohol use, including caffeine and hypnotics.

Polysomnography is necessary for the diagnosis of several disorders such as sleep apnea, narcolepsy, and periodic limb movement disorder (PLMD). A conventional polysomnogram performed in a clinical sleep laboratory allows measurement of sleep stages, respiratory effort and airflow, oxygen saturation, limb movements, heart rhythm, and additional parameters. A home sleep test usually focuses on just respiratory measures and is helpful in patients with a moderate to high likelihood of having obstructive sleep apnea. The multiple sleep latency test (MSLT) is used to measure a patient's propensity to sleep during the day and can provide crucial evidence for diagnosing narcolepsy and some other causes of sleepiness. The maintenance of wakefulness test is used to measure a patient's ability to sustain wakefulness during the daytime and can provide important evidence for evaluating the efficacy of therapies for improving sleepiness in conditions such as narcolepsy and obstructive sleep apnea.

EVALUATION OF DAYTIME SLEEPINESS

Up to 25% of the adult population has persistent daytime sleepiness that impairs an individual's ability to perform optimally in school, at work, while driving, and in other conditions that require alertness. Sleepy students often have trouble staying alert and performing well in school, and sleepy adults struggle to stay awake and focused on their work. More than half of Americans have fallen asleep while driving. An estimated 1.2 million motor vehicle crashes per year are due to drowsy drivers, causing about 20% of all serious crash injuries and deaths. One need not fall asleep to have a motor vehicle crash, as the inattention and slowed responses of drowsy drivers are major contributors. Twenty-four hours of continuous wakefulness impairs reaction time as much as a blood alcohol concentration of 0.10 g/dL (which is legally drunk in all 50 states).

Identifying and quantifying sleepiness can be challenging. First, patients may describe themselves as "sleepy," "fatigued," or "tired," and the meanings of these words may differ between patients. For clinical purposes, it is best to use the term "sleepiness" to describe a propensity to fall asleep, whereas "fatigue" is best used to describe a feeling of low physical or mental energy but without a tendency to actually sleep. Sleepiness is usually most evident when the patient is sedentary, whereas fatigue may interfere with more active pursuits. Sleepiness generally occurs with disorders that reduce the quality or quantity of sleep or that interfere with the neural mechanisms of arousal, whereas fatigue is more common in inflammatory disorders such as cancer, multiple sclerosis **(Chap. 444)**, fibromyalgia **(Chap. 373)**, chronic fatigue syndrome **(Chap. 450)**, or endocrine deficiencies such as hypothyroidism **(Chap. 383)** or Addison's disease **(Chap. 386)**. Second, sleepiness can affect judgment in a manner analogous to ethanol, such that patients may have limited insight into the condition and the extent of their functional impairment. Finally, patients may be reluctant to admit that sleepiness is a problem because they may have become unfamiliar with feeling fully alert, and because sleepiness is sometimes viewed pejoratively as

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 31-1

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

which the patient is sedentary, whereas fatigue may interfere with more active pursuits. Sleepiness generally occurs with disorders that reduce the patient's ability to maintain wakefulness. Patients may have limited insight into the condition and the extent of their functional impairment. Finally, patients may be reluctant to admit that sleepiness is a problem because they may have become unfamiliar with feeling fully alert, and because sleepiness is sometimes viewed pejoratively as reflecting poor motivation or bad sleep habits.

Table 31-1 outlines the diagnostic and therapeutic approach to the patient with a complaint of excessive daytime sleepiness.

TABLE 31-1

Evaluation of the Patient with Excessive Daytime Sleepiness

FINDINGS ON HISTORY AND PHYSICAL EXAMINATION	DIAGNOSTIC EVALUATION	DIAGNOSIS	THERAPY
Difficulty waking in the morning, rebound sleep on weekends and vacations with improvement in sleepiness	Sleep log	Insufficient sleep	Sleep education and behavioral modification to increase amount of sleep
Obesity, snoring, hypertension	Polysomnogram or home sleep test	Obstructive sleep apnea **(Chap. 297)**	Continuous positive airway pressure; upper airway surgery (e.g., uvulopalatopharyngoplasty); dental appliance; weight loss
Cataplexy, hypnagogic hallucinations, sleep paralysis	Polysomnogram and multiple sleep latency test	Narcolepsy	Stimulants (e.g., modafinil, methylphenidate); REM sleep-suppressing antidepressants (e.g., venlafaxine); pitolisant; solriamfetol; sodium oxybate
Restless legs, kicking movements during sleep	Assessment for predisposing medical conditions (e.g., iron deficiency or renal failure)	Restless legs syndrome with or without periodic limb movements	Treatment of predisposing condition; dopamine agonists (e.g., pramipexole, ropinirole); gabapentin; pregabalin; opiates
Sedating medications, stimulant withdrawal, head trauma, systemic inflammation, Parkinson's disease and other neurodegenerative disorders, hypothyroidism, encephalopathy	Thorough medical history and examination including detailed neurologic examination	Sleepiness due to a drug or medical condition	Change medications, treat underlying condition, consider stimulants

To determine the extent and impact of sleepiness on daytime function, it is helpful to ask patients about the occurrence of sleep episodes during normal waking hours, both intentional and unintentional. Specific areas to be addressed include the occurrence of inadvertent sleep episodes while driving or in other safety-related settings, sleepiness while at work or school (and its impact on performance), and the effect of sleepiness on social and family life. Standardized questionnaires such as the Epworth Sleepiness Scale are often used clinically to measure sleepiness.

Eliciting a history of daytime sleepiness is usually adequate, but objective quantification is sometimes necessary. The MSLT measures a patient's propensity to sleep under quiet conditions. An overnight polysomnogram should precede the MSLT to establish that the patient has had an adequate amount of good-quality nighttime sleep. The MSLT consists of five 20-min nap opportunities every 2 h across the day. The patient is instructed to try to fall asleep, and the major endpoints are the average latency to sleep and the occurrence of REM sleep during the naps. An average sleep latency across the naps of <8 min is considered objective evidence of excessive daytime sleepiness. REM sleep normally occurs only during nighttime sleep, and the occurrence of REM sleep in two or more of the MSLT daytime naps provides support for the diagnosis of narcolepsy.

For the safety of the individual and the general public, physicians have a responsibility to help manage issues around driving in patients with sleepiness. Legal reporting requirements vary between states and countries, but at a minimum, physicians should inform sleepy patients about their

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

INSUFFICIENT SLEEP

Insufficient sleep is probably the most common cause of excessive daytime sleepiness. The average adult needs 7.5–8 h of sleep, but on weeknights the

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

sleepiness. Legal reporting requirements vary between states and countries, but at a minimum, physicians should inform sleepy patients about their increased risk of having an accident and advise such patients not to drive a motor vehicle until the sleepiness has been treated effectively. This discussion is especially important for commercial drivers, and it should be documented in the patient's medical record.

INSUFFICIENT SLEEP

Insufficient sleep is probably the most common cause of excessive daytime sleepiness. The average adult needs 7.5–8 h of sleep, but on weeknights the average U.S. adult gets only 6.75 h of sleep. Only 30% of the U.S. adult population reports consistently obtaining sufficient sleep. Insufficient sleep is especially common among shift workers, individuals working multiple jobs, and people in lower socioeconomic groups. Most teenagers need ≥9 h of sleep, but many fail to get enough sleep because of circadian phase delay, plus social pressures to stay up late coupled with early school start times. Late evening light exposure, television viewing, video-gaming, social media, texting, and smartphone use often delay bedtimes, despite the fixed early wake times required for work or school. As is typical with any disorder that causes sleepiness, individuals with chronically insufficient sleep may feel inattentive, irritable, unmotivated, and depressed, and have difficulty with school, work, and driving. Individuals differ in their optimal amount of sleep, and it can be helpful to ask how much sleep the patient obtains on a quiet vacation when he or she can sleep without restrictions. Some patients may think that a short amount of sleep is normal or advantageous, and they may not appreciate their biological need for more sleep, especially if coffee and other stimulants mask the sleepiness. A 2-week sleep log documenting the timing of sleep and daily level of alertness is diagnostically useful and provides helpful feedback for the patient. Extending sleep to the optimal amount on a regular basis can resolve the sleepiness and other symptoms. As with any lifestyle change, extending sleep requires commitment and adjustments, but the improvements in daytime alertness make this change worthwhile.

SLEEP APNEA SYNDROMES

Respiratory dysfunction during sleep is a common, serious cause of excessive daytime sleepiness as well as of disturbed nocturnal sleep. At least 24% of middle-aged men and 9% of middle-aged women in the United States have a reduction or cessation of breathing dozens or more times each night during sleep, with 9% of men and 4% of women doing so more than a hundred times per night. These episodes may be due to an occlusion of the airway (*obstructive sleep apnea*), absence of respiratory effort (*central sleep apnea*), or a combination of these factors. Failure to recognize and treat these conditions appropriately may reduce daytime alertness and increase the risk of sleep-related motor vehicle crashes, depression, hypertension, myocardial infarction, diabetes, stroke, and mortality. Sleep apnea is particularly prevalent in overweight men and in the elderly, yet it is estimated to go undiagnosed in most affected individuals. This is unfortunate because several effective treatments are available. **Readers are referred to Chap. 297 for a comprehensive review of the diagnosis and treatment of patients with sleep apnea.**

NARCOLEPSY

Narcolepsy is characterized by difficulty sustaining wakefulness, poor regulation of REM sleep, and disturbed nocturnal sleep. All patients with narcolepsy have excessive daytime sleepiness. This sleepiness is usually moderate to severe, and in contrast to patients with disrupted sleep (e.g., sleep apnea), people with narcolepsy usually feel well rested upon awakening and then feel tired throughout much of the day. They may fall asleep at inappropriate times, but then feel refreshed again after a nap. In addition, they often experience symptoms related to an intrusion of REM sleep characteristics into wakefulness. REM sleep is characterized by dreaming and muscle paralysis, and people with narcolepsy can have: (1) sudden muscle weakness without a loss of consciousness, which is usually triggered by strong emotions (cataplexy; **Video 31-1**); (2) dream-like hallucinations at sleep onset (hypnagogic hallucinations) or upon awakening (hypnopompic hallucinations); and (3) muscle paralysis upon awakening (sleep paralysis). With severe cataplexy, an individual may be laughing at a joke and then suddenly collapse to the ground, immobile but awake for 1–2 min. With milder episodes, patients may have partial weakness of the face or neck. Narcolepsy is one of the more common causes of chronic sleepiness and affects about 1 in 2000 people in the United States. Narcolepsy typically begins between age 10 and 20; once established, the disease persists for life.

Video 31-01. Classic Cataplexy



Play Video

Narcolepsy is caused by loss of the hypothalamic neurons that produce the orexin neuropeptides (also known as hypocretins). Research in mice and dogs first demonstrated that a loss of orexin signaling due to null mutations of either the orexin neuropeptides or one of the orexin receptors causes

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

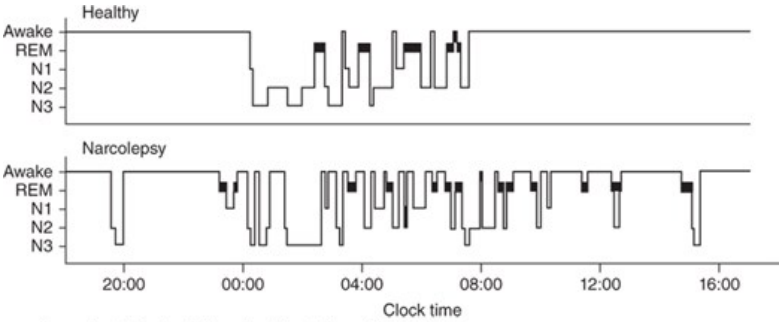
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

long episodes of wakefulness and suppress REM sleep, and thus loss of orexin signaling results in frequent intrusions of sleep during the usual waking episode, with REM sleep and fragments of REM sleep at any time of day **(Fig. 31-3)**. Patients with narcolepsy but no cataplexy (type 2 narcolepsy)

Narcolepsy is caused by loss of the hypothalamic neurons that produce the orexin neuropeptides (also known as hypocretins). Research in mice and dogs first demonstrated that a loss of orexin signaling due to null mutations of either the orexin neuropeptides or one of the orexin receptors causes sleepiness and cataplexy nearly identical to that seen in people with narcolepsy. Although genetic mutations rarely cause human narcolepsy, researchers soon discovered that patients with narcolepsy with cataplexy (now called type 1 narcolepsy) have very low or undetectable levels of orexins in their CSF, and autopsy studies showed a nearly complete loss of the orexin-producing neurons in the hypothalamus. The orexins normally promote long episodes of wakefulness and suppress REM sleep, and thus loss of orexin signaling results in frequent intrusions of sleep during the usual waking episode, with REM sleep and fragments of REM sleep at any time of day **(Fig. 31-3)**. Patients with narcolepsy but no cataplexy (type 2 narcolepsy) usually have normal orexin levels and may have other yet uncharacterized causes of their excessive daytime sleepiness.

FIGURE 31-3

Polysomnographic recordings of a healthy individual and a patient with narcolepsy. The healthy individual has a long period or NREM sleep before entering REM sleep, but the individual with narcolepsy enters rapid eye movement (REM) sleep quickly at night and has moderately fragmented sleep. During the day, the healthy subject stays awake from 8:00 A.M. until midnight, but the patient with narcolepsy dozes off frequently, with many daytime naps that include REM sleep.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Extensive evidence suggests that an autoimmune process likely causes this selective loss of the orexin-producing neurons. Certain human leukocyte antigens (HLAs) can increase the risk of autoimmune disorders **(Chap. 350)**, and narcolepsy has the strongest known HLA association. HLA DQB1*06:02 is found in >90% of people with type 1 narcolepsy, whereas it occurs in only 12–25% of the general population. Researchers now hypothesize that in people with DQB1*06:02, an immune response against influenza, *Streptococcus*, or other infections may also damage the orexin-producing neurons through a process of molecular mimicry. This mechanism may account for the eight- to twelvefold increase in new cases of narcolepsy among children in Europe who received a particular brand of H1N1 influenza A vaccine (Pandemrix). In support of this hypothesis, people with type 1 narcolepsy have heightened T cell responses against orexin peptides.

On rare occasions, narcolepsy can occur with neurologic disorders such as tumors or strokes that directly damage the orexin-producing neurons in the hypothalamus or their projections.

Diagnosis

Narcolepsy is most commonly diagnosed by the history of chronic sleepiness plus cataplexy or other symptoms. Many disorders can cause feelings of weakness, but with true cataplexy patients will describe definite functional weakness (e.g., slurred speech, dropping a cup, slumping into a chair) that has consistent emotional triggers such as laughing at a joke, happy surprise at unexpectedly seeing a friend, or intense anger. Cataplexy occurs in about half of all narcolepsy patients and is diagnostically very helpful because it occurs in almost no other disorder. In contrast, occasional hypnagogic hallucinations and sleep paralysis occur in about 20% of the general population, and these symptoms are not as diagnostically specific.

When narcolepsy is suspected, the diagnosis should be firmly established with a polysomnogram followed the next day by an MSLT. The polysomnogram helps rule out other possible causes of sleepiness such as sleep apnea and establishes that the patient had adequate sleep the night before, and the MSLT provides essential, objective evidence of sleepiness plus REM sleep dysregulation. Across the five naps of the MSLT, most patients with narcolepsy will fall asleep in <8 min on average, and they will have episodes of REM sleep in at least two of the naps. Abnormal regulation of REM sleep is also manifested by the appearance of REM sleep within 15 min of sleep onset at night, which is rare in healthy individuals sleeping at their habitual bedtime. Stimulants should be stopped 1 week before the MSLT and antidepressants should be stopped 3 weeks prior, because these medications can affect the MSLT. In addition, patients should be encouraged to obtain a fully adequate amount of sleep each night for the week prior

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

before, and the MSLT provides essential, objective evidence of sleepiness plus REM sleep dysregulation. Across the five naps of the MSLT, most medications can affect the MSLT. In addition, patients should be encouraged to obtain a fully adequate amount of sleep each night for the week prior to the test to eliminate any effects of insufficient sleep.

TREATMENT

Narcolepsy

The treatment of narcolepsy is symptomatic. Most patients with narcolepsy feel more alert after sleep, and they should be encouraged to get adequate sleep each night and to take a 15- to 20-min nap in the afternoon. This nap may be sufficient for some patients with mild narcolepsy, but most also require treatment with wake-promoting medications. Modafinil is often used because it has fewer side effects than amphetamines and a relatively long half-life; for most patients, 200–400 mg each morning is very effective. Methylphenidate (10–20 mg bid) or dextroamphetamine (10 mg bid) are also effective, but sympathomimetic side effects, anxiety, and the potential for abuse can be concerns. These medications are available in slow-release formulations, extending their duration of action and allowing easier dosing. Solriamfetol, a norepinephrine–dopamine reuptake inhibitor (75–150 mg daily), and pitolisant, a selective histamine 3 (H_3) receptor antagonist (8.9–35.6 mg daily), also improve sleepiness and have relatively few side effects. Sodium oxybate (gamma hydroxybutyrate), given at bedtime and 3–4 h later, is often very valuable in improving alertness, but it can produce excessive sedation, nausea, and confusion.

Cataplexy is usually much improved with antidepressants that increase noradrenergic or serotonergic tone because these neurotransmitters strongly suppress REM sleep and cataplexy. Venlafaxine (37.5–150 mg each morning) and fluoxetine (10–40 mg each morning) are often quite effective. The tricyclic antidepressants, such as protriptyline (10–40 mg/d) or clomipramine (25–50 mg/d) are potent suppressors of cataplexy, but their anticholinergic effects, including sedation and dry mouth, make them less attractive.[1] Sodium oxybate, twice each night, is also very helpful in reducing cataplexy.

[1]No antidepressant has been approved by the US Food and Drug Administration (FDA) for treating narcolepsy.

EVALUATION OF INSOMNIA

Insomnia is the complaint of poor sleep and usually presents as difficulty initiating or maintaining sleep. People with insomnia are dissatisfied with their sleep and feel that it impairs their ability to function well in work, school, and social situations. Affected individuals often experience fatigue, decreased mood, irritability, malaise, and cognitive impairment.

Chronic insomnia, lasting >3 months, occurs in about 10% of adults and is more common in women, older adults, people of lower socioeconomic status, and individuals with medical, psychiatric, and substance abuse disorders. Acute or short-term insomnia affects over 30% of adults and is often precipitated by stressful life events such as a major illness or loss, change of occupation, medications, and substance abuse. If the acute insomnia triggers maladaptive behaviors such as increased nocturnal light exposure, frequently checking the clock, or attempting to sleep more by napping, it can lead to chronic insomnia.

Most insomnia begins in adulthood, but many patients may be predisposed and report easily disturbed sleep predating the insomnia, suggesting that their sleep is lighter than usual. Clinical studies and animal models indicate that insomnia is associated with activation during sleep of brain areas normally active only during wakefulness. The polysomnogram is rarely used in the evaluation of insomnia, as it typically confirms the patient's subjective report of long latency to sleep and numerous awakenings but usually adds little new information. Many patients with insomnia have increased fast (beta) activity in the EEG during sleep; this fast activity is normally present only during wakefulness, which may explain why some patients report feeling awake for much of the night. The MSLT is rarely used in the evaluation of insomnia because, despite their feelings of low energy, most people with insomnia do not easily fall asleep during the day, and on the MSLT, their average sleep latencies are usually longer than normal.

Many factors can contribute to insomnia, and obtaining a careful history is essential so one can select therapies targeting the underlying factors. The assessment should focus on identifying predisposing, precipitating, and perpetuating factors.

Psychophysiological Factors

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

bedroom, as it lacks the negative associations.

Many patients with insomnia have negative expectations and conditioned arousal that interfere with sleep. These individuals may worry about their insomnia during the day and have increasing anxiety as bedtime approaches if they anticipate a poor night of sleep. While attempting to sleep, they may frequently check the clock, which only heightens anxiety and frustration. They may find it easier to sleep in a new environment rather than their bedroom, as it lacks the negative associations.

Inadequate Sleep Hygiene

Patients with insomnia sometimes develop counterproductive behaviors that contribute to their insomnia. These can include daytime napping that reduces sleep drive at night; an irregular sleep-wake schedule that disrupts their circadian rhythms; use of wake-promoting substances (e.g., caffeine, tobacco) too close to bedtime; engaging in alerting or stressful activities close to bedtime (e.g., arguing with a partner, work-related emailing and texting while in bed, sleeping with a smartphone or tablet at the bedside); and routinely using the bedroom for activities other than sleep or sex (e.g., email, television, work), so the bedroom becomes associated with arousing or stressful feelings.

Psychiatric Conditions

About 80% of patients with psychiatric disorders have sleep complaints, and about half of all chronic insomnia occurs in association with a psychiatric disorder. Depression is classically associated with early morning awakening, but it can also interfere with the onset and maintenance of sleep. Mania and hypomania can disrupt sleep and often are associated with substantial reductions in the total amount of sleep. Anxiety disorders can lead to racing thoughts and rumination that interfere with sleep and can be very problematic if the patient's mind becomes active midway through the night. Panic attacks can arise from sleep and need to be distinguished from other parasomnias. Insomnia is common in schizophrenia and other psychoses, often resulting in fragmented sleep, less deep NREM sleep, and sometimes reversal of the day-night sleep pattern.

Medications and Drugs of Abuse

A wide variety of psychoactive drugs can interfere with sleep. Caffeine, which has a half-life of 6–9 h, can disrupt sleep for up to 8–14 h, depending on the dose, variations in metabolism, and an individual's caffeine sensitivity. Insomnia can also result from use of prescription medications too close to bedtime (e.g., antidepressants, stimulants, glucocorticoids, theophylline). Conversely, withdrawal of sedating medications such as alcohol, narcotics, or benzodiazepines can cause insomnia. Alcohol taken just before bed can shorten sleep latency, but it often produces rebound insomnia 2–3 h later as it wears off. This same problem with sleep maintenance can occur with short-acting medications such as alprazolam or zolpidem.

Medical Conditions

A large number of medical conditions disrupt sleep. Pain from rheumatologic disorders or a painful neuropathy commonly disrupts sleep. Some patients may sleep poorly because of respiratory conditions such as asthma, chronic obstructive pulmonary disease, cystic fibrosis, congestive heart failure, or restrictive lung disease, and some of these disorders are worse at night due to circadian variations in airway resistance and postural changes in bed that can result in nocturnal dyspnea. Many women experience poor sleep with the hormonal changes of menopause. Gastroesophageal reflux is also a common cause of difficulty sleeping.

Neurologic Disorders

Dementia **(Chap. 29)** is often associated with poor sleep, probably due to a variety of factors, including napping during the day, altered circadian rhythms, and perhaps a weakened output of the brain's sleep-promoting mechanisms. In fact, insomnia and nighttime wandering are some of the most common causes for institutionalization of patients with dementia, because they place a larger burden on caregivers. Conversely, in cognitively intact elderly men, fragmented sleep and poor sleep quality are associated with subsequent cognitive decline. Patients with Parkinson's disease may sleep poorly due to rigidity, dementia, and other factors. Fatal familial insomnia is a very rare neurodegenerative condition caused by mutations in the prion protein gene **(Chap. 438)**, and although insomnia is a common early symptom, most patients present with other obvious neurologic signs such as dementia, myoclonus, dysarthria, or autonomic dysfunction.

TABLE 31-2

Methods to Improve Sleep Hygiene in Insomnia Patients

• If you cannot sleep within 20 min, get out of bed and read or do other relaxing activities in dim	physiology, including:

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 31-2

Methods to Improve Sleep Hygiene in Insomnia Patients

HELPFUL BEHAVIORS	BEHAVIORS TO AVOID
Use the bed only for sleep and sex • If you cannot sleep within 20 min, get out of bed and read or do other relaxing activities in dim light before returning to bed	Avoid behaviors that interfere with sleep physiology, including: • Napping, especially after 3:00 PM • Attempting to sleep too early • Caffeine after lunchtime
Make quality sleep a priority • Go to bed and get up at the same time each day • Ensure a restful environment (comfortable bed, bedroom quiet and dark)	In the 2–3 h before bedtime, avoid: • Heavy eating • Smoking or alcohol • Vigorous exercise
Develop a consistent bedtime routine. For example: • Prepare for sleep with 20–30 min of relaxation (e.g., soft music, meditation, yoga, pleasant reading) • Take a warm bath	When trying to fall asleep, avoid: • Solving problems • Thinking about life issues • Reviewing events of the day

RESTLESS LEGS SYNDROME

Patients with restless legs syndrome (RLS) report an irresistible urge to move the legs. Many patients report a creepy-crawly or unpleasant deep ache within the thighs or calves, and those with more severe RLS may have discomfort in the arms as well. For most patients with RLS, these dysesthesias and restlessness are much worse in the evening and first half of the night. The symptoms appear with inactivity and can make sitting still in an airplane or when watching a movie a miserable experience. The sensations are temporarily relieved by movement, stretching, or massage. This nocturnal discomfort usually interferes with sleep, and patients may report daytime sleepiness as a consequence. RLS is very common, affecting 5–10% of adults, and is more common in women and older adults.

A variety of factors can cause RLS. Iron deficiency is the most common treatable cause, and iron replacement should be considered if the ferritin level is <75 ng/mL. RLS can also occur with peripheral neuropathies and uremia and can be worsened by pregnancy, caffeine, alcohol, antidepressants, lithium, neuroleptics, and antihistamines. Genetic factors contribute to RLS, and polymorphisms in a variety of genes (*BTBD9, MEIS1, MAP2K5/LBXCOR*, and *PTPRD*) have been linked to RLS, although as yet, the mechanism through which they cause RLS remains unknown. Roughly one-third of patients (particularly those with an early age of onset) have multiple affected family members.

RLS is treated by addressing the underlying cause such as iron deficiency if present. Otherwise, treatment is symptomatic, and dopamine agonists or alpha-2-delta calcium channel ligands are used most frequently. Agonists of dopamine $D_{2/3}$ receptors such as pramipexole (0.25–0.5 mg q7PM) or ropinirole (0.5–4 mg q7PM) are usually quite effective, but about 25% of patients taking dopamine agonists develop augmentation, a worsening of RLS such that symptoms begin earlier in the day and can spread to other body regions. Other possible side effects of dopamine agonists include nausea, morning sedation, and increases in rewarding behaviors such as sex and gambling. Alpha-2-delta calcium channel ligands such as gabapentin (300–600 mg q7PM) and pregabalin (150–450 mg q7PM) can also be quite effective; these are less likely to cause augmentation, and they can be especially helpful in patients with concomitant pain, neuropathy, or anxiety. Opioids and benzodiazepines may also be of therapeutic value. Most patients with restless legs also experience PLMD, although the reverse is not the case.

PERIODIC LIMB MOVEMENT DISORDER

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

polysomnogram that includes recordings of the anterior tibialis and sometimes other muscles. The EEG shows that the movements of PLMD frequently cause brief arousals that disrupt sleep and can cause insomnia and daytime sleepiness. PLMD can be caused by the same factors that cause RLS (see

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

PERIODIC LIMB MOVEMENT DISORDER

PLMD involves rhythmic twitches of the legs that disrupt sleep. The movements resemble a triple flexion reflex with extensions of the great toe and dorsiflexion of the foot for 0.5–5.0 s, which recur every 20–40 s during NREM sleep, in episodes lasting from minutes to hours. PLMD is diagnosed by a polysomnogram that includes recordings of the anterior tibialis and sometimes other muscles. The EEG shows that the movements of PLMD frequently cause brief arousals that disrupt sleep and can cause insomnia and daytime sleepiness. PLMD can be caused by the same factors that cause RLS (see above), and the frequency of leg movements improves with the same medications used for RLS, including dopamine agonists. Genetic studies identified polymorphisms associated with both RLS and PLMD, suggesting that they may have a common pathophysiology.

PARASOMNIAS

Parasomnias are abnormal behaviors or experiences that arise from or occur during sleep. A variety of parasomnias can occur during NREM sleep, from brief confusional arousals to sleepwalking and night terrors. The presenting complaint is usually related to the behavior itself, but the parasomnias can disturb sleep continuity or lead to mild impairments in daytime alertness. Two main parasomnias occur in REM sleep: REM sleep behavior disorder (RBD) and nightmares.

Sleepwalking (Somnambulism)

Patients affected by this disorder carry out automatic motor activities that range from simple to complex. Individuals may walk, urinate inappropriately, eat, exit the house, or drive a car with minimal awareness. It may be difficult to arouse the patient to wakefulness, and some individuals may respond to attempted awakening with agitation or violence. In general, it is safest to lead the patient back to bed, at which point he or she will often fall back asleep. Sleepwalking arises from NREM stage N3 sleep, usually in the first few hours of the night, and the EEG initially shows the slow cortical activity of deep NREM sleep even when the patient is moving about. Sleepwalking is most common in children and adolescents, when deep NREM sleep is most abundant. About 15% of children have occasional sleepwalking, and it persists in about 1% of adults. Episodes are usually isolated but may be recurrent in 1–6% of patients. The cause is unknown, although it has a familial basis in roughly one-third of cases. Sleepwalking can be worsened by stress, alcohol, and insufficient sleep, which subsequently causes an increase in deep NREM sleep. These should be addressed if present. Small studies have shown some efficacy of antidepressants and benzodiazepines; relaxation techniques and hypnosis can also be helpful. Patients and their families should improve home safety (e.g., replace glass doors, remove low tables to avoid tripping) to minimize the chance of injury if sleepwalking occurs.

Sleep Terrors

This disorder occurs primarily in young children during the first few hours of sleep during NREM stage N3 sleep. The child often sits up during sleep and screams, exhibiting autonomic arousal with sweating, tachycardia, large pupils, and hyperventilation. The individual may be difficult to arouse and rarely recalls the episode on awakening in the morning. Treatment usually consists of reassuring parents that the condition is self-limited and benign, and like sleepwalking, it may improve by avoiding insufficient sleep.

Sleep Enuresis

Bedwetting, like sleepwalking and night terrors, is another parasomnia that occurs during sleep in the young. Before age 5 or 6 years, nocturnal enuresis should be considered a normal feature of development. The condition usually improves spontaneously by puberty, persists in 1–3% of adolescents, and is rare in adulthood. Treatment consists of bladder training exercises and behavioral therapy. Symptomatic pharmacotherapy is usually accomplished in adults with desmopressin (0.2 mg qhs), oxybutynin chloride (5 mg qhs), or imipramine (10–25 mg qhs). Important causes of nocturnal enuresis in patients who were previously continent for 6–12 months include urinary tract infections or malformations, cauda equina lesions, emotional disturbances, epilepsy, sleep apnea, and certain medications.

Sleep Bruxism

Bruxism is an involuntary, forceful grinding of teeth during sleep that affects 10–20% of the population. The patient is usually unaware of the problem. The typical age of onset is 17–20 years, and spontaneous remission usually occurs by age 40. In many cases, the diagnosis is made during dental examination, damage is minor, and no treatment is indicated. In more severe cases, treatment with a mouth guard is necessary to prevent tooth injury. Stress management, benzodiazepines, and biofeedback can be useful when bruxism is a manifestation of psychological stress.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

RBD **(Video 31-2)** is distinct from other parasomnias in that it occurs during REM sleep. The patient or the bed partner usually reports agitated or violent behavior during sleep, and upon awakening, the patient can often report a dream that matches the accompanying movements. During normal

examination, damage is minor, and no treatment is indicated. In more severe cases, treatment with a mouth guard is necessary to prevent tooth injury. Stress management, benzodiazepines, and biofeedback can be useful when bruxism is a manifestation of psychological stress.

REM Sleep Behavior Disorder (RBD)

RBD **(Video 31-2)** is distinct from other parasomnias in that it occurs during REM sleep. The patient or the bed partner usually reports agitated or violent behavior during sleep, and upon awakening, the patient can often report a dream that matches the accompanying movements. During normal REM sleep, nearly all nonrespiratory skeletal muscles are paralyzed, but in patients with RBD, dramatic limb movements such as punching or kicking lasting seconds to minutes occur during REM sleep, and it is not uncommon for the patient or the bed partner to be injured.

Video 31-02. REM Sleep Behavior Disorder



Play Video

The prevalence of RBD increases with age, afflicting about 2% of adults aged >70, and is about twice as common in men. Within 12 years of disease onset, half of RBD patients develop a synucleinopathy such as Parkinson's disease **(Chap. 435)** or dementia with Lewy bodies **(Chap. 434)**, or occasionally multiple system atrophy **(Chap. 440)**, and over 90% develop a synucleinopathy by 25 years. RBD can occur in patients taking antidepressants, and in some, these medications may unmask this early indicator of neurodegeneration. Synucleinopathies probably cause neuronal loss in brainstem regions that regulate muscle paralysis during REM sleep, and loss of these neurons permits movements to break through during REM sleep. RBD also occurs in about 30% of patients with narcolepsy, but the underlying cause is probably different, as they seem to be at no increased risk of a neurodegenerative disorder.

Many patients with RBD have sustained improvement with clonazepam (0.5–2.0 mg qhs).[3] Melatonin at doses up to 9 mg nightly may also prevent attacks.

[3]No medications have been approved by the FDA for the treatment of RBD.

CIRCADIAN RHYTHM SLEEP DISORDERS

A subset of patients presenting with either insomnia or hypersomnia may have a disorder of sleep *timing* rather than sleep *generation.* Disorders of sleep timing can be either organic (i.e., due to an abnormality of circadian pacemaker[s]) or environmental/behavioral (i.e., due to a disruption of environmental synchronizers). Effective therapies aim to entrain the circadian rhythm of sleep propensity to the appropriate behavioral phase.

Delayed Sleep-Wake Phase Disorder

DSWPD is characterized by: (1) sleep onset and wake times persistently later than desired; (2) actual sleep times at nearly the same clock hours daily; and (3) if conducted at the habitual delayed sleep time, essentially normal sleep on polysomnography (except for delayed sleep onset). About half of patients with DSWPD exhibit an abnormally delayed endogenous circadian phase, which can be assessed by measuring the onset of secretion of melatonin in either the blood or saliva; this is best done in a dimly lit environment as light suppresses melatonin secretion. Dim-light melatonin onset (DLMO) in DSWPD patients occurs later in the evening than normal, which is about 8:00–9:00 P.M. (i.e., about 1–2 h before habitual bedtime). Patients tend to be young adults. The delayed circadian phase could be due to: (1) an abnormally long, genetically determined intrinsic period of the endogenous circadian pacemaker; (2) reduced phase-advancing capacity of the pacemaker; (3) slower buildup of homeostatic sleep drive during wakefulness; or (4) an irregular prior sleep-wake schedule, characterized by frequent nights when the patient chooses to remain awake while exposed to artificial light well past midnight (for personal, social, school, or work reasons). In most cases, it is difficult to distinguish among these factors, as patients with either a behaviorally induced or biologically driven circadian phase delay may both exhibit a similar circadian phase delay in DLMO, and both factors make it difficult to fall asleep at the desired hour. Late onset of dim-light melatonin secretion can help distinguish DSWPD from other forms of sleep-onset insomnia. DSWPD is a chronic condition that can persist for years and may not respond to attempts to reestablish normal bedtime hours. Treatment methods involving phototherapy with blue-enriched light during the morning hours and/or melatonin administration in the evening hours show promise in these patients, although the relapse rate is high.

Advanced Sleep-Wake Phase Disorder

that they cannot sleep past 5:00 A.M., with twice that number complaining that they wake up too early at least several times per week. Patients with ASWPD are sleepy during the evening hours, even in social settings. Sleep-wake timing in ASWPD patients can interfere with a normal social life.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Advanced Sleep-Wake Phase Disorder

Advanced sleep-wake phase disorder (ASWPD) is the converse of DSWPD. Most commonly, this syndrome occurs in older people, 15% of whom report that they cannot sleep past 5:00 A.M., with twice that number complaining that they wake up too early at least several times per week. Patients with ASWPD are sleepy during the evening hours, even in social settings. Sleep-wake timing in ASWPD patients can interfere with a normal social life. Patients with this circadian rhythm sleep disorder can be distinguished from those who have early wakening due to insomnia because ASWPD patients show early onset of dim-light melatonin secretion.

In addition to age-related ASWPD, an early-onset familial variant of this condition has also been reported. In two families in which ASWPD was inherited in an autosomal dominant pattern, the syndrome was due to missense mutations in a circadian clock component (in the casein kinase binding domain of *PER2* in one family, and in casein kinase I delta in the other) that shortens the circadian period. Patients with ASWPD may benefit from bright light and/or blue enriched phototherapy during the evening hours to reset the circadian pacemaker to a later hour.

Non-24-h Sleep-Wake Rhythm Disorder

Non-24-h sleep-wake rhythm disorder (N24SWD) most commonly occurs when the primary synchronizing input (i.e., the light-dark cycle) from the environment to the circadian pacemaker is lost (as occurs in many blind people with no light perception), and the maximal phase-advancing capacity of the circadian pacemaker in response to nonphotic cues cannot accommodate the difference between the 24-h geophysical day and the intrinsic period of the patient's circadian pacemaker, resulting in loss of entrainment to the 24-h day. The sleep of most blind patients with N24SWD is restricted to the nighttime hours due to social or occupational demands. Despite this regular sleep-wake schedule, affected patients with N24SWD are nonetheless unable to maintain a stable phase relationship between the output of the non-entrained circadian pacemaker and the 24-h day. Therefore, most blind patients present with intermittent bouts of insomnia. When the blind patient's endogenous circadian rhythms are out of phase with the local environment, nighttime insomnia coexists with excessive daytime sleepiness. Conversely, when the endogenous circadian rhythms of those same patients are in phase with the local environment, symptoms remit. The interval between symptomatic phases may last several weeks to several months in blind patients with N24SWD, depending on the period of the underlying nonentrained rhythm and the 24-h day. Nightly low-dose (0.5 mg) melatonin administration may improve sleep and, in some cases, induce synchronization of the circadian pacemaker. In sighted patients, N24SWD can be caused by self-selected exposure to artificial light that inadvertently entrains the circadian pacemaker to a >24-h schedule, and these individuals present with an incremental pattern of successive delays in sleep timing, progressing in and out of phase with local time—a clinical presentation that is seldom seen in blind patients with N24SWD.

Shift-Work Disorder

More than 7 million workers in the United States regularly work at night, either on a permanent or rotating schedule. Many more begin the commute to work or school between 4:00 A.M. and 7:00 A.M., requiring them to commute and then work during a time of day that they would otherwise be asleep. In addition, each week, millions of "day" workers and students elect to remain awake at night or awaken very early in the morning to work or study to meet work or school deadlines, drive long distances, compete in sporting events, or participate in recreational activities. Such schedules can result in both sleep loss and misalignment of circadian rhythms with respect to the sleep-wake cycle.

The circadian timing system usually fails to adapt successfully to the inverted schedules required by overnight work or the phase advance required by early morning (4:00 A.M. to 7:00 A.M.) start times. This leads to a misalignment between the desired work-rest schedule and the output of the pacemaker, resulting in disturbed daytime sleep in most such individuals. Excessive work hours (per day or per week), insufficient time off between consecutive days of work or school, and frequent travel across time zones may be contributing factors. Sleep deficiency, increased length of time awake prior to work, and misalignment of circadian phase impair alertness and performance, increase reaction time, and increase risk of performance lapses, thereby resulting in greater safety hazards among night workers and other sleep-deprived individuals. Sleep disturbance nearly doubles the risk of a fatal work accident. In addition, long-term night-shift workers have higher rates of breast, colorectal, and prostate cancer and of cardiac, gastrointestinal, metabolic, and reproductive disorders. The World Health Organization has added night-shift work to its list of probable carcinogens.

Sleep onset begins in local brain regions before gradually sweeping over the entire brain as sensory thresholds rise and consciousness is lost. A sleepy individual struggling to remain awake may attempt to continue performing routine and familiar motor tasks during the transition state between wakefulness and stage N1 sleep, while unable to adequately process sensory input from the environment. Such sleep-related attentional failures typically last only seconds but are known on occasion to persist for longer durations. Motor vehicle operators who fail to heed the warning signs of

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

vehicle. There is a significant increase in the risk of sleep-related, fatal-to-the-driver highway crashes in the early morning and late afternoon hours, coincident with bimodal peaks in the daily rhythm of sleep tendency.

typically last only seconds but are known on occasion to persist for longer durations. Motor vehicle operators who fail to heed the warning signs of sleepiness are especially vulnerable to sleep-related accidents, as sleep processes can slow reaction times, induce automatic behavior, and intrude involuntarily upon the waking brain, causing catastrophic consequences—including 6400 fatalities and 50,000 debilitating injuries in the United States annually. For this reason, an expert consensus panel has concluded that individuals who have slept <2 h in the prior 24 h are unfit to drive a motor vehicle. There is a significant increase in the risk of sleep-related, fatal-to-the-driver highway crashes in the early morning and late afternoon hours, coincident with bimodal peaks in the daily rhythm of sleep tendency.

Physicians who work prolonged shifts, especially intermittent overnight shifts, constitute another group of workers at greater risk for accidents and other adverse consequences of lack of sleep and misalignment of the circadian rhythm. Recurrent scheduling of resident physicians to work shifts of ≥24 consecutive hours impairs psychomotor performance to a degree that is comparable to alcohol intoxication, doubles the risk of attentional failures among intensive care unit resident physicians working at night, and significantly increases the risk of serious medical errors in intensive care units, including a fivefold increase in the risk of serious diagnostic mistakes. Some 20% of hospital resident physicians report making a fatigue-related mistake that injured a patient, and 5% admit making a fatigue-related mistake that resulted in the death of a patient. Moreover, working for >24 consecutive hours increases the risk of percutaneous injuries and more than doubles the risk of motor vehicle crashes during the commute home. For these reasons, in 2008, the National Academy of Medicine concluded that the practice of scheduling resident physicians to work for >16 consecutive hours without sleep is hazardous for both resident physicians and their patients.

Of individuals scheduled to work at night or in the early morning hours, 5–15% have much greater-than-average difficulties remaining awake during night work and sleeping during the day; these individuals are diagnosed with chronic and severe shift-work disorder (SWD). Patients with this disorder have a level of excessive sleepiness during work at night or in the early morning and insomnia during day sleep that the physician judges to be clinically significant; the condition is associated with an increased risk of sleep-related accidents and with some of the illnesses associated with night-shift work. Patients with chronic and severe SWD are profoundly sleepy at work. In fact, their sleep latencies during night work average just 2 min, comparable to mean daytime sleep latency durations of patients with narcolepsy or severe sleep apnea.

TREATMENT

Shift-Work Disorder

Caffeine is frequently used by night workers to promote wakefulness. However, it cannot forestall sleep indefinitely, and it does not shield users from sleep-related performance lapses. Postural changes, exercise, and strategic placement of nap opportunities can sometimes temporarily reduce the risk of fatigue-related performance lapses. Properly timed exposure to blue-enriched light or bright white light can directly enhance alertness and facilitate more rapid adaptation to night-shift work.

Modafinil (200 mg) or armodafinil (150 mg) 30–60 min before the start of an 8-h overnight shift is an effective treatment for the excessive sleepiness during night work in patients with SWD. Although treatment with modafinil or armodafinil significantly improves performance and reduces sleep propensity and the risk of lapses of attention during night work, affected patients remain excessively sleepy.

Fatigue risk management programs for night-shift workers should promote education about sleep, increase awareness of the hazards associated with sleep deficiency and night work, and screen for common sleep disorders. Work schedules should be designed to minimize: (1) exposure to night work; (2) the frequency of shift rotations; (3) the number of consecutive night shifts; and (4) the duration of night shifts.

Jet Lag Disorder

Each year, >60 million people fly from one time zone to another, often resulting in excessive daytime sleepiness, sleep-onset insomnia, and frequent arousals from sleep, particularly in the latter half of the night. The syndrome is transient, typically lasting 2–14 d depending on the number of time zones crossed, the direction of travel, and the traveler's age and phase-shifting capacity. Travelers who spend more time outdoors at their destination reportedly adapt more quickly than those who remain in hotel or seminar rooms, presumably due to brighter (outdoor) light exposure. Avoidance of antecedent sleep loss or napping on the afternoon prior to overnight travel can reduce the difficulties associated with extended wakefulness. Laboratory studies suggest that low doses of melatonin can enhance sleep efficiency, but only if taken when endogenous melatonin concentrations are low (i.e., during the biologic daytime).

In addition to jet lag associated with travel across time zones, many patients report a behavioral pattern that has been termed *social jet lag*, in which

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

MEDICAL IMPLICATIONS OF CIRCADIAN RHYTHMICITY

In addition to jet lag associated with travel across time zones, many patients report a behavioral pattern that has been termed *social jet lag*, in which bedtimes and wake times on weekends or days off occur 4–8 h later than during the week. Such recurrent displacement of the timing of the sleep-wake cycle is common in adolescents and young adults and is associated with delayed circadian phase, sleep-onset insomnia, excessive daytime sleepiness, poorer academic performance, and increased risk of both obesity and depressive symptoms.

MEDICAL IMPLICATIONS OF CIRCADIAN RHYTHMICITY

Prominent circadian variations have been reported in the incidence of acute myocardial infarction, sudden cardiac death, and stroke, the leading causes of death in the United States. Platelet aggregability is increased in the early morning hours, coincident with the peak incidence of these cardiovascular events. Recurrent circadian disruption combined with chronic sleep deficiency, such as occurs during night-shift work, is associated with increased plasma glucose concentrations after a meal due to inadequate pancreatic insulin secretion. Night-shift workers with elevated fasting glucose have an increased risk of progressing to diabetes. Blood pressure of night workers with sleep apnea is higher than that of day workers. A better understanding of the possible role of circadian rhythmicity in the acute destabilization of a chronic condition such as atherosclerotic disease could improve the understanding of its pathophysiology.

Diagnostic and therapeutic procedures may also be affected by the time of day at which data are collected. Examples include blood pressure, body temperature, the dexamethasone suppression test, and plasma cortisol levels. The timing of chemotherapy administration has been reported to have an effect on the outcome of treatment. In addition, both the toxicity and effectiveness of drugs can vary with time of day. For example, more than a fivefold difference has been observed in mortality rates after administration of toxic agents to experimental animals at different times of day. Anesthetic agents are particularly sensitive to time-of-day effects. Finally, the physician must be aware of the public health risks associated with the ever-increasing demands made by the 24/7 schedules in our round-the-clock society.

ACKNOWLEDGMENT

John W. Winkelman, MD, PhD, and Gary S. Richardson, MD, contributed to this chapter in prior editions, and some material from their work has been retained here.

FURTHER READING

Cash RE et al: Association between sleep duration and ideal cardiovascular health among US adults, National Health and Nutrition Examination Survey. Prev Chronic Dis 17:E43, 2020. [PubMed: 32530395]

Chinoy ED et al: Unrestricted evening use of light-emitting tablet computers delays self-selected bedtime and disrupts circadian timing and alertness. Physiol Rep 6:e13692, 2018. [PubMed: 29845764]

Fultz NE et al: Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep. Science 366:628, 2019. [PubMed: 31672896]

Holth JK et al: The sleep-wake cycle regulates brain interstitial fluid tau in mice and CSF tau in humans. Science 363:880, 2019. [PubMed: 30679382]

Landrigan CP et al: Effect on patient safety of a resident physician schedule without 24-hour shifts. N Engl J Med 382:2514, 2020. [PubMed: 32579812]

Lee ML et al: High risk of near-crash driving events following night-shift work. Proc Natl Acad Sci USA 113:176, 2016. [PubMed: 26699470]

Lim AS et al: Sleep is related to neuron numbers in the ventrolateral preoptic/intermediate nucleus in older adults with and without Alzheimer's disease. Brain 137:2847, 2014. [PubMed: 25142380]

Mcalpine CS et al: Sleep modulates haematopoiesis and protects against atherosclerosis. Nature 566:383, 2019. [PubMed: 30760925]

Riemann D et al: The neurobiology, investigation, and treatment of chronic insomnia. Lancet Neurol 14:547, 2015. [PubMed: 25895933]

Scammell TE: Narcolepsy. N Engl J Med 373:2654, 2015. [PubMed: 26716917]

Sletten TL et al: Efficacy of melatonin with behavioural sleep-wake scheduling for delayed sleep-wake phase disorder: a double-blind, randomised clinical trial. PLoS Med 15:e1002587, 2018. [PubMed: 29912983]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Scammell TE: Narcolepsy. N Engl J Med 373:2654, 2015. [PubMed: 26716917]

Scammell TE et al: Neural circuitry of wakefulness and sleep. Neuron 93:747, 2017. [PubMed: 28231463]

Sletten TL et al: Efficacy of melatonin with behavioural sleep-wake scheduling for delayed sleep-wake phase disorder: a double-blind, randomised clinical trial. PLoS Med 15:e1002587, 2018. [PubMed: 29912983]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 31: Sleep Disorders, Thomas E. Scammell; Clifford B. Saper; Charles A. Czeisler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 32: Disorders of the Eye

Jonathan C. Horton

THE HUMAN VISUAL SYSTEM

The visual system provides a supremely efficient means for the rapid assimilation of information from the environment to aid in the guidance of behavior. The act of seeing begins with the capture of images focused by the cornea and lens on a light-sensitive membrane in the back of the eye called the *retina*. The retina is actually part of the brain, banished to the periphery to serve as a transducer for the conversion of patterns of light energy into neuronal signals. Light is absorbed by pigment in two types of photoreceptors: rods and cones. In the human retina, there are 100 million rods and 5 million cones. The rods operate in dim (scotopic) illumination. The cones function under daylight (photopic) conditions. The cone system is specialized for color perception and high spatial resolution. The majority of cones are within the macula, the portion of the retina that serves the central 10° of vision. In the middle of the macula, a small pit termed the *fovea*, packed exclusively with cones, provides the best visual acuity.

Photoreceptors hyperpolarize in response to light, activating bipolar, amacrine, and horizontal cells in the inner nuclear layer. After processing of photoreceptor responses by this complex retinal circuit, the flow of sensory information ultimately converges on a final common pathway: the ganglion cells. These cells translate the visual image impinging on the retina into a continuously varying barrage of action potentials that propagates along the primary optic pathway to visual centers within the brain. There are a million ganglion cells in each retina and hence a million fibers in each optic nerve.

Ganglion cell axons sweep along the inner surface of the retina in the nerve fiber layer, exit the eye at the optic disc, and travel through the optic nerve, optic chiasm, and optic tract to reach targets in the brain. The majority of fibers synapse on cells in the lateral geniculate body, a thalamic relay station. Cells in the lateral geniculate body project in turn to the primary visual cortex. This afferent retinogeniculocortical sensory pathway provides the neural substrate for visual perception. Although the lateral geniculate body is the main target of the retina, separate classes of ganglion cells project to other subcortical visual nuclei involved in different functions. Ganglion cells that mediate pupillary constriction and circadian rhythms are light sensitive owing to a novel visual pigment, melanopsin. Pupil responses are mediated by input to the pretectal olivary nuclei in the midbrain. The pretectal nuclei send their output to the Edinger-Westphal nuclei, which in turn provide parasympathetic innervation to the iris sphincter via an interneuron in the ciliary ganglion. Circadian rhythms are timed by a retinal projection to the suprachiasmatic nucleus. Visual orientation and eye movements are served by retinal input to the superior colliculus. Gaze stabilization and optokinetic reflexes are governed by a group of small retinal targets known collectively as the *brainstem accessory optic system*.

The eyes must be rotated constantly within their orbits to place and maintain targets of visual interest on the fovea. This activity, called *foveation*, or looking, is governed by an elaborate efferent motor system. Each eye is moved by six extraocular muscles that are supplied by cranial nerves from the oculomotor (III), trochlear (IV), and abducens (VI) nuclei. Activity in these ocular motor nuclei is coordinated by pontine and midbrain mechanisms for smooth pursuit, saccades, and gaze stabilization during head and body movements. Large regions of the frontal and parietooccipital cortex control these brainstem eye movement centers by providing descending supranuclear input.

CLINICAL ASSESSMENT OF VISUAL FUNCTION

REFRACTIVE STATE

In approaching a patient with reduced vision, the first step is to decide whether refractive error is responsible. In *emmetropia*, parallel rays from infinity are focused perfectly on the retina. Sadly, this condition is enjoyed by only a minority of the population. In *myopia*, the globe is too long, and light rays come to a focal point in front of the retina. Near objects can be seen clearly, but distant objects require a diverging lens in front of the eye. In *hyperopia*, the globe is too short, and hence, a converging lens is used to supplement the refractive power of the eye. In *astigmatism*, the corneal surface is not perfectly spherical, necessitating a cylindrical corrective lens. Most patients elect to wear eyeglasses or contact lenses to neutralize refractive error. An alternative is to permanently alter the refractive properties of the cornea by performing laser in situ keratomileusis (LASIK) or

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

objects. To compensate for presbyopia, an emmetropic patient must use reading glasses. A patient already wearing glasses for distance correction usually switches to bifocals. The only exception is a myopic patient, who may achieve clear vision at near simply by removing glasses containing the

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

surface is not perfectly spherical, necessitating a cylindrical corrective lens. Most patients elect to wear eyeglasses or contact lenses to neutralize refractive error. An alternative is to permanently alter the refractive properties of the cornea by performing laser in situ keratomileusis (LASIK) or photorefractive keratectomy (PRK).

With the onset of middle age, *presbyopia* develops as the lens within the eye becomes unable to increase its refractive power to accommodate on near objects. To compensate for presbyopia, an emmetropic patient must use reading glasses. A patient already wearing glasses for distance correction usually switches to bifocals. The only exception is a myopic patient, who may achieve clear vision at near simply by removing glasses containing the distance prescription.

Refractive errors usually develop slowly and remain stable after adolescence, except in unusual circumstances. For example, the acute onset of diabetes mellitus can produce sudden myopia because of lens edema induced by hyperglycemia. Testing vision through a pinhole aperture is a useful way to screen quickly for refractive error. If visual acuity is better through a pinhole than it is with the unaided eye, the patient needs refraction to obtain best corrected visual acuity.

VISUAL ACUITY

The Snellen chart is used to test acuity at a distance of 6 m (20 ft). For convenience, a scale version of the Snellen chart called the Rosenbaum card is held at 36 cm (14 in.) from the patient **(Fig. 32-1)**. All subjects should be able to read the 6/6 m (20/20 ft) line with each eye using their refractive correction, if any. Patients who need reading glasses because of presbyopia must wear them for accurate testing with the Rosenbaum card. If 6/6 (20/20) acuity is not present in each eye, the deficiency in vision must be explained. If it is worse than 6/240 (20/800), acuity should be recorded in terms of counting fingers, hand motions, light perception, or no light perception. Legal blindness is defined by the Internal Revenue Service as a best corrected acuity of 6/60 (20/200) or less in the better eye or a binocular visual field subtending 20° or less. Loss of vision in one eye only does not constitute legal blindness. For driving, the laws vary by state, but most require a corrected acuity of 6/12 (20/40) in at least one eye for unrestricted privileges. Patients who develop a homonymous hemianopia should not drive.

FIGURE 32-1

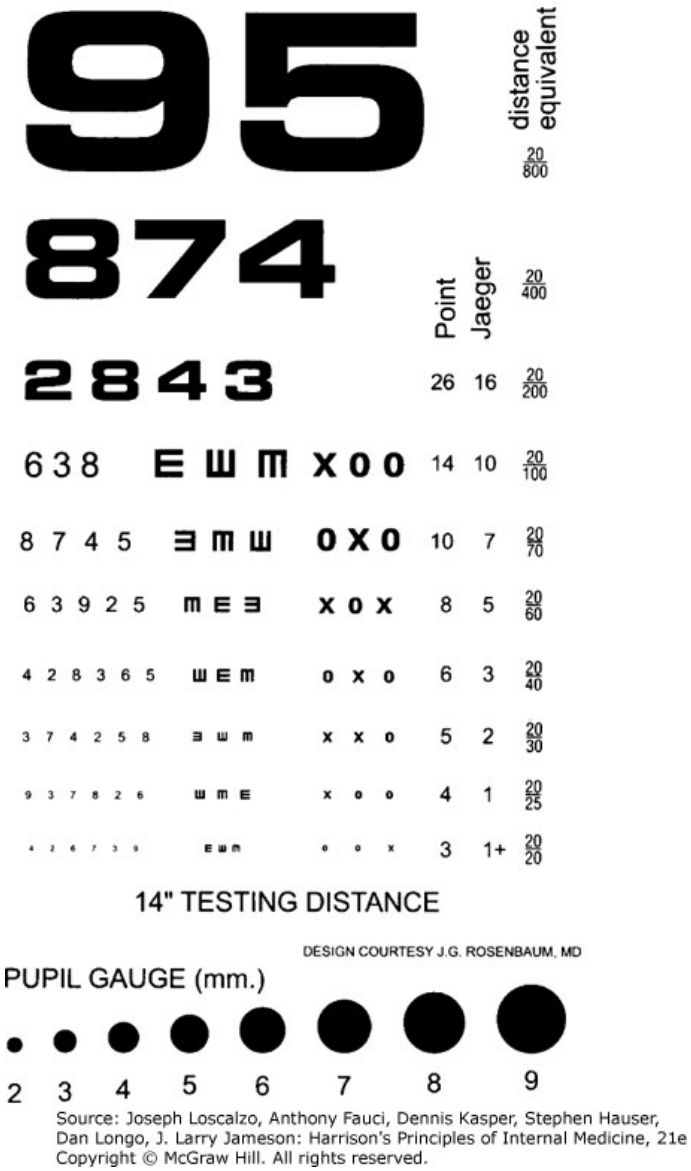
The Rosenbaum card is a miniature, scale version of the Snellen chart for testing visual acuity at near. When the visual acuity is recorded, the Snellen distance equivalent should bear a notation indicating that vision was tested at near, not at 6 m (20 ft), or else the Jaeger number system should be used to report the acuity. (*Design Courtesy J.G. Rosenbaum MD.*)



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

PUPILS

The pupils should be tested individually in dim light with the patient fixating on a distant target. There is no need to check the near response if the pupils respond briskly to light, because isolated loss of constriction (miosis) to accommodation does not occur. For this reason, the ubiquitous abbreviation PERRLA (pupils equal, round, and reactive to light and accommodation) implies a wasted effort with the last step. However, it is important to test the near response if the light response is poor or absent. Light-near dissociation occurs with neurosyphilis (Argyll Robertson pupil), with lesions of the dorsal midbrain (*Parinaud's syndrome*), and after aberrant regeneration (oculomotor nerve palsy, Adie's tonic pupil).

An eye with no light perception has no pupillary response to direct light stimulation. If the retina or optic nerve is only partially injured, the direct pupillary response will be weaker than the consensual pupillary response evoked by shining a light into the healthy fellow eye. A *relative afferent pupillary defect* (Marcus Gunn pupil) is elicited with the swinging flashlight test **(Fig. 32-2)**. It is an extremely useful sign in retrobulbar optic neuritis and other optic nerve diseases, in which it may be the sole objective evidence for disease. In bilateral optic neuropathy, no afferent pupil defect is present if the optic nerves are affected equally.

FIGURE 32-2

Demonstration of a relative afferent pupil defect (Marcus Gunn pupil) in the left eye, done with the patient fixating on a distant

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

A. Swinging the flashlight back over to the healthy right eye would result in symmetric constriction back to the appearance shown in ***B.*** Note that the pupils always remain equal; the damage to the left retina/optic nerve is revealed by weaker bilateral pupil constriction to a flashlight in the left eye

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Demonstration of a relative afferent pupil defect (Marcus Gunn pupil) in the left eye, done with the patient fixating on a distant target. ***A.*** With dim background lighting, the pupils are equal and relatively large. ***B.*** Shining a flashlight into the right eye evokes equal, strong constriction of both pupils. ***C.*** Swinging the flashlight over to the damaged left eye causes dilation of both pupils, although they remain smaller than in ***A.*** Swinging the flashlight back over to the healthy right eye would result in symmetric constriction back to the appearance shown in ***B.*** Note that the pupils always remain equal; the damage to the left retina/optic nerve is revealed by weaker bilateral pupil constriction to a flashlight in the left eye compared with the right eye. *(From P Levatin: Arch Ophthalmol 62:768, 1959. Copyright © 1959 American Medical Association. All rights reserved.)*



A



B



C

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Subtle inequality in pupil size, up to 0.5 mm, is a fairly common finding in normal persons. The diagnosis of essential or physiologic anisocoria is secure as long as the relative pupil asymmetry remains constant as ambient lighting varies. Anisocoria that increases in dim light indicates a sympathetic paresis of the iris dilator muscle. The triad of miosis with ipsilateral ptosis and anhidrosis constitutes *Horner's syndrome*, although anhidrosis is an inconstant feature. A drop of 1% apraclonidine produces no effect on the normal pupil, but the miotic pupil dilates because of denervation hypersensitivity. Brainstem stroke, carotid dissection, and neoplasm impinging on the sympathetic chain occasionally are identified as the cause of Horner's syndrome, but most cases are idiopathic.

Anisocoria that increases in bright light suggests a parasympathetic palsy. The first concern is an oculomotor nerve paresis. This possibility is excluded

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

When the near stimulus is removed, the pupil redilates very slowly compared with the normal pupil, hence the term *tonic pupil*. In *Adie's syndrome*, a tonic pupil is present, sometimes in conjunction with weak or absent tendon reflexes in the lower extremities. This benign disorder, which occurs

Anisocoria that increases in bright light suggests a parasympathetic palsy. The first concern is an oculomotor nerve paresis. This possibility is excluded if the eye movements are full and the patient has no ptosis or diplopia. Acute pupillary dilation (mydriasis) can result from damage to the ciliary ganglion in the orbit. Common mechanisms are infection (herpes zoster, influenza), trauma (blunt, penetrating, surgical), and ischemia (diabetes, temporal arteritis). After denervation of the iris sphincter, the pupil does not respond well to light, but the response to near is often relatively intact. When the near stimulus is removed, the pupil redilates very slowly compared with the normal pupil, hence the term *tonic pupil*. In *Adie's syndrome*, a tonic pupil is present, sometimes in conjunction with weak or absent tendon reflexes in the lower extremities. This benign disorder, which occurs predominantly in healthy young women, is assumed to represent a mild dysautonomia. Tonic pupils are also associated with multiple system atrophy, segmental hypohidrosis, diabetes, and amyloidosis. Occasionally, a tonic pupil is discovered incidentally in an otherwise completely normal, asymptomatic individual. The diagnosis is confirmed by placing a drop of dilute (0.125%) pilocarpine into each eye. Denervation hypersensitivity produces pupillary constriction in a tonic pupil, whereas the normal pupil shows no response. Pharmacologic dilatation from accidental or deliberate instillation of anticholinergic (atropine, scopolamine) drops can produce pupillary mydriasis. Gardener's pupil refers to mydriasis induced by exposure to tropane alkaloids, contained in plants such as deadly nightshade, jimsonweed, or angel's trumpet. When an anticholinergic agent is responsible for pupil dilation, 1% pilocarpine causes no constriction.

Both pupils are affected equally by systemic medications. They are small with narcotic use (morphine, oxycodone) and large with anticholinergics (scopolamine). Parasympathetic agents (pilocarpine) used to treat glaucoma produce miosis. In any patient with an unexplained pupillary abnormality, a slit-lamp examination is helpful to exclude surgical trauma to the iris, an occult foreign body, perforating injury, intraocular inflammation, adhesions (synechia), angle-closure glaucoma, and iris sphincter rupture from blunt trauma.

EYE MOVEMENTS AND ALIGNMENT

Eye movements are tested by asking the patient, with both eyes open, to pursue a small target such as a pen tip into the cardinal fields of gaze. Normal ocular versions are smooth, symmetric, full, and maintained in all directions without nystagmus. Saccades, or quick refixation eye movements, are assessed by having the patient look back and forth between two stationary targets. The eyes should move rapidly and accurately in a single jump to their target. Ocular alignment can be judged by holding a penlight directly in front of the patient at about 1 m. If the eyes are straight, the corneal light reflex will be centered in the middle of each pupil. To test eye alignment more precisely, the cover test is useful. The patient is instructed to look at a small fixation target in the distance. One eye is occluded with a paddle or hand, while the other eye is observed. If the viewing eye shifts position to take up fixation on the target, it was misaligned. If it remains motionless, the first eye is uncovered and the test is repeated on the second eye. If neither eye moves, the eyes are aligned orthotropically. If the eyes are orthotropic in primary gaze but the patient complains of diplopia, the cover test should be performed with the head tilted or turned in whatever direction elicits diplopia. With practice, the examiner can detect an ocular deviation (heterotropia) as small as 1–2° with the cover test. In a patient with vertical diplopia, a small deviation can be difficult to detect and easy to dismiss. The magnitude of the deviation can be measured by placing a prism in front of the misaligned eye to determine the power required to neutralize the fixation shift evoked by covering the other eye. Temporary press-on plastic Fresnel prisms, prism eyeglasses, or eye muscle surgery can be used to restore binocular alignment.

STEREOPSIS

Stereoacuity is determined by presenting targets with retinal disparity separately to each eye by using polarized images. The most popular office tests measure a range of thresholds from 800 to 40 s of arc. Normal stereoacuity is 40 s of arc. If a patient achieves this level of stereoacuity, one is assured that the eyes are aligned orthotropically and that vision is intact in each eye. Random dot stereograms have no monocular depth cues and provide an excellent screening test for strabismus.

COLOR VISION

The retina contains three classes of cones, with visual pigments of differing peak spectral sensitivity: red (560 nm), green (530 nm), and blue (430 nm). The red and green cone pigments are encoded on the X chromosome, and the blue cone pigment on chromosome 7. Mutations of the blue cone pigment are exceedingly rare. Mutations of the red and green pigments cause congenital X-linked color blindness in 8% of males. Affected individuals are not truly color blind; rather, they differ from normal subjects in the way they perceive color and how they combine primary monochromatic lights to match a particular color. Anomalous trichromats have three cone types, but a mutation in one cone pigment (usually red or green) causes a shift in peak spectral sensitivity, altering the proportion of primary colors required to achieve a color match. Dichromats have only two cone types and therefore will accept a color match based on only two primary colors. Anomalous trichromats and dichromats have 6/6 (20/20) visual acuity, but their hue discrimination is impaired. Ishihara color plates can be used to detect red-green color blindness. The test plates contain a hidden number that is

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The Ishihara plates often are used to detect acquired defects in color vision, although they are intended as a screening test for congenital color blindness. Acquired defects in color vision frequently result from disease of the macula or optic nerve. For example, patients with a history of optic

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

hue discrimination is impaired. Ishihara color plates can be used to detect red-green color blindness. The test plates contain a hidden number that is visible only to subjects with color confusion from red-green color blindness. Because color blindness is almost exclusively X-linked, it is worthwhile screening only male children.

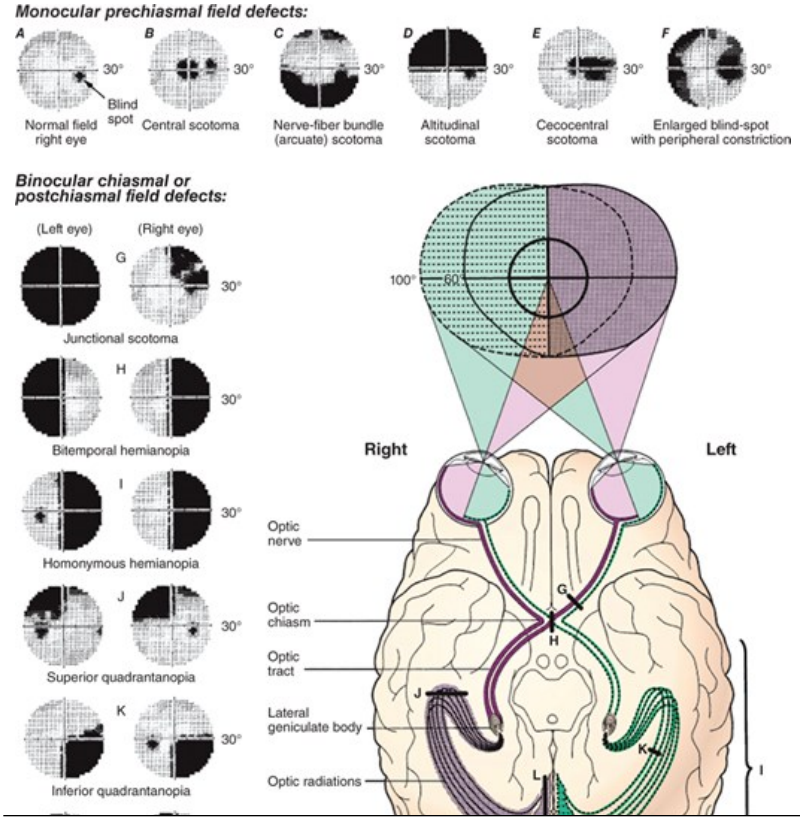
The Ishihara plates often are used to detect acquired defects in color vision, although they are intended as a screening test for congenital color blindness. Acquired defects in color vision frequently result from disease of the macula or optic nerve. For example, patients with a history of optic neuritis often complain of color desaturation long after their visual acuity has returned to normal. Color blindness also can result from bilateral strokes involving the ventral portion of the occipital lobe (cerebral achromatopsia). Such patients can perceive only shades of gray and also may have difficulty recognizing faces (prosopagnosia) **(Chap. 30)**. Infarcts of the dominant occipital lobe sometimes give rise to color anomia. Affected patients can discriminate colors but cannot name them.

VISUAL FIELDS

Vision can be impaired by damage to the visual system anywhere from the eyes to the occipital lobes. One can localize the site of the lesion with considerable accuracy by mapping the visual field deficit by finger confrontation and then correlating it with the topographic anatomy of the visual pathway **(Fig. 32-3)**. Quantitative visual field mapping is performed by computer-driven perimeters that present a target of variable intensity at fixed positions in the visual field **(Fig. 32-3A)**. By generating an automated printout of light thresholds, these static perimeters provide a sensitive means of detecting scotomas in the visual field. They are exceedingly useful for serial assessment of visual function in chronic diseases such as glaucoma and pseudotumor cerebri.

FIGURE 32-3

Ventral view of the brain, correlating patterns of visual field loss with the sites of lesions in the visual pathway. The visual fields overlap partially, creating 120° of central binocular field flanked by a 40° monocular crescent on either side. The visual field maps in this figure were done with a computer-driven perimeter (Humphrey Instruments, Carl Zeiss, Inc.). It plots the retinal sensitivity to light in the central 30° by using a gray scale format. Areas of visual field loss are shown in black. The examples of common monocular, prechiasmal field defects are all shown for the right eye. By convention, the visual fields are always recorded with the left eye's field on the left and the right eye's field on the right, just as the patient sees the world.



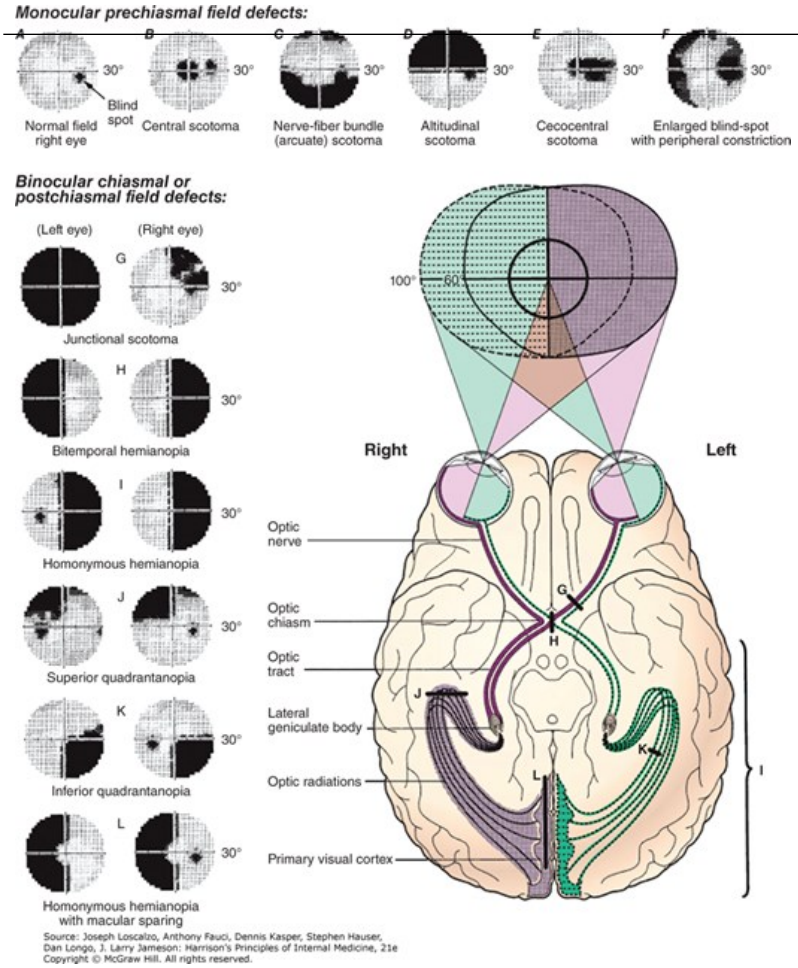
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The crux of visual field analysis is to decide whether a lesion is before, at, or behind the optic chiasm. If a scotoma is confined to one eye, it must be due

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



The crux of visual field analysis is to decide whether a lesion is before, at, or behind the optic chiasm. If a scotoma is confined to one eye, it must be due to a lesion anterior to the chiasm, involving either the optic nerve or the retina. Retinal lesions produce scotomas that correspond optically to their location in the fundus. For example, a superior-nasal retinal detachment results in an inferior-temporal field cut. Damage to the macula causes a central scotoma **(Fig. 32-3B)**.

Optic nerve disease produces characteristic patterns of visual field loss. Glaucoma selectively destroys axons that enter the superotemporal or inferotemporal poles of the optic disc, resulting in arcuate scotomas shaped like a Turkish scimitar, which emanate from the blind spot and curve around fixation to end flat against the horizontal meridian **(Fig. 32-3C)**. This type of field defect mirrors the arrangement of the nerve fiber layer in the temporal retina. Arcuate or nerve fiber layer scotomas also result from optic neuritis, ischemic optic neuropathy, optic disc drusen, and branch retinal artery or vein occlusion.

Damage to the entire upper or lower pole of the optic disc causes an altitudinal field cut that follows the horizontal meridian **(Fig. 32-3D)**. This pattern of visual field loss is typical of ischemic optic neuropathy but also results from retinal vascular occlusion, advanced glaucoma, and optic neuritis.

About half the fibers in the optic nerve originate from ganglion cells serving the macula. Damage to papillomacular fibers causes a cecocentral scotoma that encompasses the blind spot and macula **(Fig. 32-3E)**. If the damage is irreversible, pallor eventually appears in the temporal portion of the optic disc. Temporal pallor from a cecocentral scotoma may develop in optic neuritis, nutritional optic neuropathy, toxic optic neuropathy, Leber's hereditary optic neuropathy, Kjer's dominant optic atrophy, and compressive optic neuropathy. It is worth mentioning that the temporal side of the optic disc is slightly paler than the nasal side in most normal individuals. Therefore, it sometimes can be difficult to decide whether the temporal pallor visible on fundus examination represents a pathologic change. Pallor of the nasal rim of the optic disc is a less equivocal sign of optic atrophy.

At the optic chiasm, fibers from nasal ganglion cells decussate into the contralateral optic tract. Crossed fibers are damaged more by compression than are uncrossed fibers. As a result, mass lesions of the sellar region cause a temporal hemianopia in each eye. Tumors anterior to the optic chiasm, such as meningiomas of the tuberculum sella, produce a junctional scotoma characterized by an optic neuropathy in one eye and a superior-temporal field cut in the other eye **(Fig. 32-3G)**. More symmetric compression of the optic chiasm by a pituitary adenoma **(see Fig. 380-1)**, meningioma

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

It is difficult to localize a postchiasmal lesion accurately, because injury anywhere in the optic tract, lateral geniculate body, optic radiations, or visual cortex can produce a homonymous hemianopia (i.e., a temporal hemifield defect in the contralateral eye and a matching nasal hemifield defect in the

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

visible on fundus examination represents a pathologic change. Pallor of the nasal rim of the optic disc is a less equivocal sign of optic atrophy.

as meningiomas of the tuberculum sella, produce a junctional scotoma characterized by an optic neuropathy in one eye and a superior-temporal field cut in the other eye **(Fig. 32-3G)**. More symmetric compression of the optic chiasm by a pituitary adenoma **(see Fig. 380-1)**, meningioma, craniopharyngioma, glioma, or aneurysm results in a bitemporal hemianopia **(Fig. 32-3H)**. The insidious development of a bitemporal hemianopia often goes unnoticed by the patient and will escape detection by the physician unless each eye is tested separately.

It is difficult to localize a postchiasmal lesion accurately, because injury anywhere in the optic tract, lateral geniculate body, optic radiations, or visual cortex can produce a homonymous hemianopia (i.e., a temporal hemifield defect in the contralateral eye and a matching nasal hemifield defect in the ipsilateral eye) **(Fig. 32-3I)**. A unilateral postchiasmal lesion leaves the visual acuity in each eye unaffected, although the patient may read the letters on only the left or right half of the eye chart. Lesions of the optic radiations tend to cause poorly matched or incongruous field defects in each eye. Damage to the optic radiations in the temporal lobe (Meyer's loop) produces a superior quadrantic homonymous hemianopia **(Fig. 32-3J)**, whereas injury to the optic radiations in the parietal lobe results in an inferior quadrantic homonymous hemianopia **(Fig. 32-3K)**. Lesions of the primary visual cortex give rise to dense, congruous hemianopic field defects. Occlusion of the posterior cerebral artery supplying the occipital lobe is a common cause of total homonymous hemianopia. Some patients have macular sparing, because the central field representation at the tip of the occipital lobe is supplied by collaterals from the middle cerebral artery **(Fig. 32-3L)**. Destruction of both occipital lobes produces cortical blindness. This condition can be distinguished from bilateral prechiasmal visual loss by noting that the pupil responses and optic fundi remain normal.

Partial recovery of homonymous hemianopia has been reported through computer-based rehabilitation therapy. During daily training sessions, patients fixate a central target while visual stimuli are presented within the blind region. The premise of vision restoration programs is that extra stimulation can promote recovery of partially damaged tissue located at the fringe of a cortical lesion. When fixation is controlled rigorously, however, no improvement of the visual fields can be demonstrated. No effective treatment exists for homonymous hemianopia caused by permanent brain damage.

DISORDERS

RED OR PAINFUL EYE

Corneal Abrasions

Corneal abrasions are seen best by placing a drop of fluorescein in the eye and looking with the slit lamp, using a cobalt-blue light. A penlight with a blue filter will suffice if a slit lamp is not available. Damage to the corneal epithelium is revealed by yellow fluorescence of the basement membrane exposed by loss of the overlying epithelium. It is important to check for foreign bodies. To search the conjunctival fornices, the lower lid should be pulled down and the upper lid everted. A foreign body can be removed with a moistened cotton-tipped applicator after a drop of a topical anesthetic such as proparacaine has been placed in the eye. Alternatively, it may be possible to flush the foreign body from the eye by irrigating copiously with saline or artificial tears. If the corneal epithelium has been abraded, antibiotic ointment and a patch may be applied to the eye. A drop of an intermediate-acting cycloplegic such as cyclopentolate hydrochloride 1% helps reduce pain by relaxing the ciliary body. The eye should be reexamined the next day. Minor abrasions may not require patching, antibiotics, or cycloplegia.

Subconjunctival Hemorrhage

This results from rupture of small vessels bridging the potential space between the episclera and the conjunctiva. Blood dissecting into this space can produce a spectacular red eye, but vision is not affected and the hemorrhage resolves without treatment. Subconjunctival hemorrhage is usually spontaneous but can result from blunt trauma, eye rubbing, or vigorous coughing. Occasionally, it is a clue to an underlying bleeding disorder.

Pinguecula

Pinguecula is a small, raised conjunctival nodule, usually at the nasal limbus. In adults, such lesions are extremely common and have little significance unless they become inflamed (pingueculitis). They are more apt to occur in workers with outdoor exposure. A *pterygium* resembles a pinguecula but has crossed the limbus to encroach on the corneal surface. Removal is justified when symptoms of irritation or blurring develop, but recurrence is common.

Blepharitis

This refers to inflammation of the eyelids. The most common form occurs in association with acne rosacea or seborrheic dermatitis. The eyelid

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

hordeolum occurs after suppurative infection of the oil-secreting meibomian glands within the tarsal plate of the eyelid. Topical antibiotics such as bacitracin/polymyxin B ophthalmic ointment can be applied. Systemic antibiotics, usually tetracyclines or azithromycin, sometimes are necessary for

This refers to inflammation of the eyelids. The most common form occurs in association with acne rosacea or seborrheic dermatitis. The eyelid margins usually are colonized heavily by staphylococci. Upon close inspection, they appear greasy, ulcerated, and crusted with scaling debris that clings to the lashes. Treatment consists of strict eyelid hygiene, applying warm compresses, and eyelash scrubs with baby shampoo. An external *hordeolum* (sty) is caused by staphylococcal infection of the superficial accessory glands of Zeis or Moll located in the eyelid margins. An internal hordeolum occurs after suppurative infection of the oil-secreting meibomian glands within the tarsal plate of the eyelid. Topical antibiotics such as bacitracin/polymyxin B ophthalmic ointment can be applied. Systemic antibiotics, usually tetracyclines or azithromycin, sometimes are necessary for treatment of meibomian gland inflammation (meibomitis) or chronic, severe blepharitis. A *chalazion* is a painless, chronic granulomatous inflammation of a meibomian gland that produces a pealike nodule within the eyelid. It can be incised and drained, but injection with glucocorticoids is equally effective. Basal cell, squamous cell, or meibomian gland carcinoma should be suspected with any nonhealing ulcerative lesion of the eyelids.

Dacryocystitis

An inflammation of the lacrimal drainage system, dacryocystitis can produce epiphora (tearing) and ocular injection. Gentle pressure over the lacrimal sac evokes pain and reflux of mucus or pus from the tear puncta. Dacryocystitis usually occurs after obstruction of the lacrimal system. It is treated with topical and systemic antibiotics, followed by probing, silicone stent intubation, or surgery to reestablish patency. *Entropion* (inversion of the eyelid) or *ectropion* (sagging or eversion of the eyelid) can also lead to epiphora and ocular irritation.

Conjunctivitis

Conjunctivitis is the most common cause of a red, irritated eye. Pain is minimal, and visual acuity is reduced only slightly. The most common viral etiology is adenovirus infection. It causes a watery discharge, a mild foreign-body sensation, and photophobia. Bacterial infection tends to produce a more mucopurulent exudate. Mild cases of infectious conjunctivitis usually are treated empirically with broad-spectrum topical ocular antibiotics such as sulfacetamide 10%, polymyxin-bacitracin, or a trimethoprim-polymyxin combination. Smears and cultures usually are reserved for severe, resistant, or recurrent cases of conjunctivitis. To prevent contagion, patients should be admonished to wash their hands frequently, not to touch their eyes, and to avoid direct contact with others.

Allergic Conjunctivitis

This condition is extremely common and often is mistaken for infectious conjunctivitis. Itching, redness, and epiphora are typical. The palpebral conjunctiva may become hypertropic with giant excrescences called cobblestone papillae. Irritation from contact lenses or any chronic foreign body also can induce formation of cobblestone papillae. *Atopic conjunctivitis* occurs in subjects with atopic dermatitis or asthma. Symptoms caused by allergic conjunctivitis can be alleviated with cold compresses, topical vasoconstrictors, antihistamines (olopatadine), and mast cell stabilizers (cromolyn). Topical glucocorticoid solutions provide dramatic relief of immune-mediated forms of conjunctivitis, but their long-term use is ill advised because of the complications of glaucoma, cataract, and secondary infection. Topical nonsteroidal anti-inflammatory drugs (NSAIDs; ketorolac) are better alternatives.

Keratoconjunctivitis Sicca

Also known as dry eye, this produces a burning foreign-body sensation, injection, and photophobia. In mild cases, the eye appears surprisingly normal, but tear production measured by wetting of a filter paper (Schirmer strip) is deficient. A variety of systemic drugs, including antihistaminic, anticholinergic, and psychotropic medications, result in dry eye by reducing lacrimal secretion. Disorders that involve the lacrimal gland directly, such as sarcoidosis and Sjögren's syndrome, also cause dry eye. Patients may develop dry eye after radiation therapy if the treatment field includes the orbits. Problems with ocular drying are also common after lesions affecting cranial nerve V or VII. Corneal anesthesia is particularly dangerous, because the absence of a normal blink reflex exposes the cornea to injury without pain to warn the patient. Dry eye is managed by frequent and liberal application of artificial tears and ocular lubricants. In severe cases, the tear puncta can be plugged or cauterized to reduce lacrimal outflow.

Keratitis

Keratitis is a threat to vision because of the risk of corneal clouding, scarring, and perforation. Worldwide, the two leading causes of blindness from keratitis are trachoma from chlamydial infection and vitamin A deficiency related to malnutrition. In the United States, contact lenses play a major role in corneal infection and ulceration. They should not be worn by anyone with an active eye infection. In evaluating the cornea, it is important to differentiate between a superficial infection (*keratoconjunctivitis*) and a deeper, more serious ulcerative process. The latter is accompanied by greater

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Giemsa stain, potassium hydroxide (KOH) prep, and cultures. Fortified topical antibiotics are most effective, supplemented with subconjunctival antibiotics as required. A fungal etiology should always be considered in a patient with keratitis. Fungal infection is common in warm humid climates

differentiate between a superficial infection (*keratoconjunctivitis*) and a deeper, more serious ulcerative process. The latter is accompanied by greater visual loss, pain, photophobia, redness, and discharge. Slit-lamp examination shows disruption of the corneal epithelium, a cloudy infiltrate or abscess in the stroma, and an inflammatory cellular reaction in the anterior chamber. In severe cases, pus settles at the bottom of the anterior chamber, giving rise to a hypopyon. Immediate empirical antibiotic therapy should be initiated after corneal scrapings are obtained for Gram's stain, Giemsa stain, potassium hydroxide (KOH) prep, and cultures. Fortified topical antibiotics are most effective, supplemented with subconjunctival antibiotics as required. A fungal etiology should always be considered in a patient with keratitis. Fungal infection is common in warm humid climates, especially after penetration of the cornea by plant or vegetable material. Acanthamoeba keratitis is associated with improper disinfection of contact lenses.

Herpes Simplex

The *herpesviruses* are a major cause of blindness from keratitis. Most adults in the United States have serum antibodies to herpes simplex, indicating prior viral infection **(Chap. 192)**. Primary ocular infection generally is caused by herpes simplex type 1 rather than type 2. It manifests as a unilateral follicular blepharoconjunctivitis that is easily confused with adenoviral conjunctivitis, unless telltale vesicles are present on the eyelids or conjunctiva. Recurrent ocular infection arises from reactivation of latent herpesvirus. A dendritic pattern of corneal epithelial ulceration revealed by fluorescein staining is pathognomonic for herpes infection but often not present. Involvement of both eyes is extremely rare. Corneal stromal inflammation produces edema, vascularization, and iridocyclitis. Herpes keratitis is treated with cycloplegia and either a topical antiviral (trifluridine, ganciclovir) or an oral antiviral (acyclovir, valacyclovir) agent. Topical glucocorticoids are effective in mitigating corneal scarring but generally are reserved for cases involving stromal damage. Risks include corneal melting, perforation, prolonged infection, and glaucoma.

Herpes Zoster

Herpes zoster from reactivation of latent varicella (chickenpox) virus causes a dermatomal pattern of painful vesicular dermatitis **(Chap. 193)**. Ocular symptoms can occur after zoster eruption in any branch of the trigeminal nerve but are particularly common when vesicles form on the nose, reflecting nasociliary (V1) nerve involvement (Hutchinson's sign). Herpes zoster ophthalmicus produces corneal dendrites, which can be difficult to distinguish from those seen in herpes simplex. Stromal keratitis, anterior uveitis, raised intraocular pressure, ocular motor nerve palsies, acute retinal necrosis, and postherpetic scarring and neuralgia are other common sequelae. Herpes zoster ophthalmicus is treated with antiviral agents and cycloplegics. In severe cases, glucocorticoids may be added to prevent permanent visual loss from corneal scarring. Shingles should be prevented by vaccination of all healthy adults aged 50 years and older.

Episcleritis

This is an inflammation of the episclera, a thin layer of connective tissue between the conjunctiva and the sclera. Episcleritis resembles conjunctivitis, but it is a more localized process and discharge is absent. Most cases of episcleritis are idiopathic, but some occur in the setting of an autoimmune disease. *Scleritis* refers to a deeper, more severe inflammatory process that frequently is associated with a connective tissue disease such as rheumatoid arthritis, lupus erythematosus, polyarteritis nodosa, granulomatosis with polyangiitis, or relapsing polychondritis. The inflammation and thickening of the sclera can be diffuse or nodular. In anterior forms of scleritis, the globe assumes a violet hue and the patient complains of severe ocular tenderness and pain. With posterior scleritis, the pain and redness may be less marked, but there is often proptosis, choroidal effusion, reduced motility, and visual loss. Episcleritis and scleritis should be treated with NSAIDs. If these agents fail, topical or even systemic glucocorticoid therapy may be necessary, especially if an underlying autoimmune process is active.

Anterior Uveitis

Involving the anterior structures of the eye, uveitis was previously called *iritis* or *iridocyclitis.* The diagnosis requires slit-lamp examination to identify inflammatory cells floating in the aqueous humor or deposited on the corneal endothelium (keratic precipitates). Anterior uveitis develops in sarcoidosis, ankylosing spondylitis, juvenile idiopathic arthritis, inflammatory bowel disease, psoriasis, reactive arthritis, and Behçet's disease. It also is associated with herpes infections, syphilis, Lyme disease, onchocerciasis, tuberculosis, and leprosy. Although anterior uveitis can occur in conjunction with many diseases, no cause is found to explain the majority of cases. For this reason, laboratory evaluation usually is reserved for patients with recurrent or severe anterior uveitis. Treatment is aimed at reducing inflammation and scarring by judicious use of topical glucocorticoids. Dilatation of the pupil reduces pain and prevents the formation of synechiae.

Posterior Uveitis

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

eye. Posterior uveitis is a manifestation of autoimmune diseases such as sarcoidosis, Behçet's disease, Vogt-Koyanagi-Harada syndrome, and inflammatory bowel disease. It also accompanies diseases such as toxoplasmosis, onchocerciasis, cysticercosis, coccidioidomycosis, toxocariasis, and

Posterior Uveitis

This diagnosis is made by observing inflammation of the vitreous, retina, or choroid on fundus examination. It is more likely than anterior uveitis to be associated with an identifiable systemic disease. Some patients have panuveitis, or inflammation of both the anterior and posterior segments of the eye. Posterior uveitis is a manifestation of autoimmune diseases such as sarcoidosis, Behçet's disease, Vogt-Koyanagi-Harada syndrome, and inflammatory bowel disease. It also accompanies diseases such as toxoplasmosis, onchocerciasis, cysticercosis, coccidioidomycosis, toxocariasis, and histoplasmosis; infections caused by organisms such as *Candida*, *Pneumocystis carinii*, *Cryptococcus*, *Aspergillus*, herpes, and cytomegalovirus **(see Fig. 195-1)**; and other diseases, such as syphilis, Lyme disease, tuberculosis, cat-scratch disease, Whipple's disease, and brucellosis. In multiple sclerosis, chronic inflammatory changes can develop in the extreme periphery of the retina (pars planitis or intermediate uveitis). Glucocorticoids have been the mainstay of treatment for noninfectious uveitis. Biologic agents that target proinflammatory cytokines, such as the tumor necrosis factor alpha (TNF-α) inhibitor adalimumab, are effective at preventing vision loss in chronic uveitis.

Acute Angle-Closure Glaucoma

This is an unusual but frequently misdiagnosed cause of a red, painful eye. Asian populations have a particularly high risk of angle-closure glaucoma. Susceptible eyes have a shallow anterior chamber because the eye has either a short axial length (hyperopia) or a lens enlarged by the gradual development of cataract. When the pupil becomes mid-dilated, the peripheral iris blocks aqueous outflow via the anterior chamber angle and the intraocular pressure rises abruptly, producing pain, injection, corneal edema, obscurations, and blurred vision. In some patients, ocular symptoms are overshadowed by nausea, vomiting, or headache, prompting a fruitless workup for abdominal or neurologic disease. The diagnosis is made by measuring the intraocular pressure during an acute attack or by performing gonioscopy, a procedure that allows one to observe a narrow chamber angle with a mirrored contact lens. Acute angle closure is treated with acetazolamide (PO or IV), topical beta blockers, prostaglandin analogues, α_2-adrenergic agonists, and pilocarpine to induce miosis. If these measures fail, a laser can be used to create a hole in the peripheral iris to relieve pupillary block. Many physicians are reluctant to dilate patients routinely for fundus examination because they fear precipitating an angle-closure glaucoma. The risk is actually remote and more than outweighed by the potential benefit to patients of discovering a hidden fundus lesion visible only through a fully dilated pupil. Moreover, a single attack of angle closure after pharmacologic dilatation rarely causes any permanent damage to the eye and serves as an inadvertent provocative test to identify patients with narrow angles who would benefit from prophylactic laser iridectomy.

Endophthalmitis

This results from bacterial, viral, fungal, or parasitic infection of the internal structures of the eye. It usually is acquired by hematogenous seeding from a remote site. Chronically ill, diabetic, or immunosuppressed patients, especially those with a history of indwelling IV catheters or positive blood cultures, are at greatest risk for endogenous endophthalmitis. Although most patients have ocular pain and injection, visual loss is sometimes the only symptom. Septic emboli from a diseased heart valve or a dental abscess that lodge in the retinal circulation can give rise to endophthalmitis. White-centered retinal hemorrhages known as Roth's spots **(Fig. 32-4)** are considered pathognomonic for subacute bacterial endocarditis, but they also appear in leukemia, diabetes, and many other conditions. Endophthalmitis occurs as a complication of ocular surgery, especially glaucoma filtering, occasionally months or even years after the operation. An occult penetrating foreign body or unrecognized trauma to the globe should be considered in any patient with unexplained intraocular infection or inflammation.

FIGURE 32-4

Roth's spot, cotton-wool spot, and retinal hemorrhages in a 48-year-old liver transplant patient with candidemia from immunosuppression. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

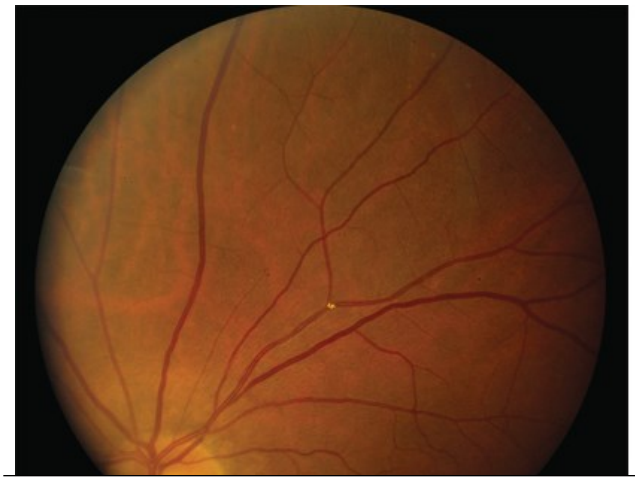
TRANSIENT OR SUDDEN VISUAL LOSS

Amaurosis Fugax

This term refers to a transient ischemic attack of the retina **(Chap. 427)**. Because neural tissue has a high rate of metabolism, interruption of blood flow to the retina for more than a few seconds results in *transient monocular blindness*, a term used interchangeably with amaurosis fugax. Patients describe a rapid fading of vision like a curtain descending, sometimes affecting only a portion of the visual field. Amaurosis fugax usually results from an embolus that becomes stuck within a retinal arteriole **(Fig. 32-5)**. If the embolus breaks up or passes, flow is restored and vision returns quickly to normal without permanent damage. With prolonged interruption of blood flow, the inner retina suffers infarction. Ophthalmoscopy reveals zones of whitened, edematous retina following the distribution of branch retinal arterioles. Complete occlusion of the central retinal artery produces arrest of blood flow and a milky retina with a cherry-red fovea **(Fig. 32-6)**. Emboli are composed of cholesterol (Hollenhorst plaque), calcium, or platelet-fibrin debris. The most common source is an atherosclerotic plaque in the carotid artery or aorta, although emboli also can arise from the heart, especially in patients with diseased valves, atrial fibrillation, or wall motion abnormalities.

FIGURE 32-5

Hollenhorst plaque lodged at the bifurcation of a retinal arteriole proves that a patient is shedding emboli from the carotid artery, great vessels, or heart. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 32-6

Central retinal artery occlusion in a 78-year-old man reducing acuity to counting fingers in the right eye. Note the splinter hemorrhage on the optic disc and the slightly milky appearance to the macula with a cherry-red fovea. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

In rare instances, amaurosis fugax results from low central retinal artery perfusion pressure in a patient with a critical stenosis of the ipsilateral carotid artery and poor collateral flow via the circle of Willis. In this situation, amaurosis fugax develops when there is a dip in systemic blood pressure or a slight worsening of the carotid stenosis. Sometimes there is contralateral motor or sensory loss, indicating concomitant hemispheric cerebral ischemia.

Retinal arterial occlusion also occurs rarely in association with retinal migraine, lupus erythematosus, anticardiolipin antibodies, anticoagulant deficiency states (protein S, protein C, and antithrombin deficiency), Susac's syndrome, pregnancy, IV drug abuse, blood dyscrasias, dysproteinemias, and temporal arteritis.

Marked *systemic hypertension* causes sclerosis of retinal arterioles, splinter hemorrhages, focal infarcts of the nerve fiber layer (cotton-wool spots),

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

optic disc infarction from sudden hypoperfusion.

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Marked *systemic hypertension* causes sclerosis of retinal arterioles, splinter hemorrhages, focal infarcts of the nerve fiber layer (cotton-wool spots), and leakage of lipid and fluid (hard exudate) into the macula **(Fig. 32-7)**. In hypertensive crisis, sudden visual loss can result from ischemia induced by vasospasm of retinal arterioles. In addition, visual loss can occur from ischemic optic disc swelling. Patients with acute hypertensive retinopathy should be treated by lowering the blood pressure. However, the blood pressure should not be reduced precipitously, because there is a danger of optic disc infarction from sudden hypoperfusion.

FIGURE 32-7

Hypertensive retinopathy with blurred optic disc, scattered hemorrhages, cotton-wool spots (nerve fiber layer infarcts), and foveal exudate in a 62-year-old man with chronic renal failure and a systolic blood pressure of 220. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*

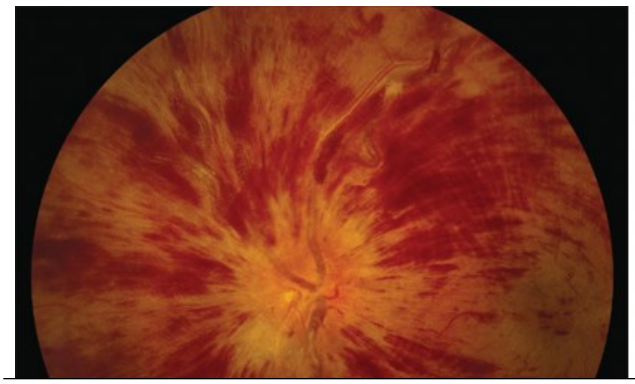


Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Impending *branch* or *central retinal vein occlusion* can produce prolonged visual obscurations that resemble those described by patients with amaurosis fugax. The veins appear engorged and phlebitic, with numerous retinal hemorrhages **(Fig. 32-8)**. In some patients, venous blood flow recovers spontaneously, whereas others evolve a frank obstruction with extensive retinal bleeding ("blood and thunder" appearance), infarction, and visual loss. Venous occlusion of the retina is often idiopathic, but hypertension, diabetes, and glaucoma are prominent risk factors. Polycythemia, thrombocythemia, or other factors leading to an underlying hypercoagulable state should be corrected; aspirin treatment may be beneficial.

FIGURE 32-8

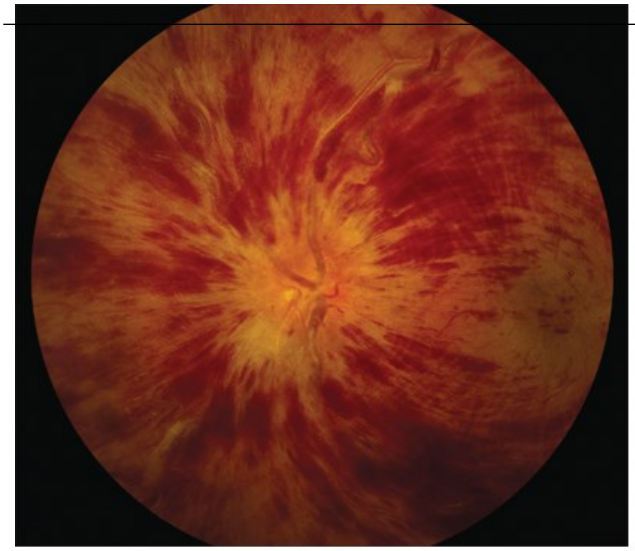
Central retinal vein occlusion can produce massive retinal hemorrhage ("blood and thunder"), ischemia, and vision loss. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Anterior Ischemic Optic Neuropathy (AION)

This is caused by insufficient blood flow through the posterior ciliary arteries that supply the optic disc. It produces painless monocular visual loss that is sudden in onset, followed sometimes by stuttering progression. The optic disc is edematous and usually bordered by nerve fiber layer splinter hemorrhages **(Fig. 32-9)**. AION is divided into two forms: arteritic and nonarteritic. The nonarteritic form is most common. No specific cause is known, although diabetes, renal failure, and hypertension are common risk factors. Case reports have linked erectile dysfunction drugs to AION, but a causal association is doubtful. Evidence is strong that a crowded disc architecture and small optic cup predispose to the development of nonarteritic AION. In patients with such a "disc-at-risk," the advent of AION in one eye increases the likelihood of the same event occurring in the other eye. No treatment is available for nonarteritic AION; glucocorticoids should not be prescribed.

FIGURE 32-9

Anterior ischemic optic neuropathy from temporal arteritis in a 64-year-old woman with acute disc swelling, splinter hemorrhages, visual loss, and an erythrocyte sedimentation rate of 60 mm/h. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

About 5% of patients, especially Caucasian females aged >60, have the arteritic form of AION in conjunction with giant cell (temporal) arteritis **(Chap**

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

symptoms of giant cell arteritis. Symptoms of polymyalgia rheumatica may be present; the sedimentation rate and C-reactive protein level are usually elevated. In a patient with visual loss from suspected arteritic AION, temporal artery biopsy is mandatory to confirm the diagnosis. Administer

ACCESS Medicine
Access Provided by: SILVERCHAIR
Copyright © McGraw Hill. All rights reserved.

About 5% of patients, especially Caucasian females aged >60, have the arteritic form of AION in conjunction with giant cell (temporal) arteritis **(Chap. 363)**. It is urgent to recognize arteritic AION so that high doses of glucocorticoids can be instituted immediately to prevent blindness in the second eye. Tocilizumab, a monoclonal antibody against interleukin 6 receptor, is an effective alternative to glucocorticoids for sustained suppression of symptoms of giant cell arteritis. Symptoms of polymyalgia rheumatica may be present; the sedimentation rate and C-reactive protein level are usually elevated. In a patient with visual loss from suspected arteritic AION, temporal artery biopsy is mandatory to confirm the diagnosis. Administer glucocorticoids immediately, without waiting for the biopsy to be completed. The biopsy should be obtained as soon as practical, because prolonged glucocorticoid treatment can hide inflammatory changes. It is important to harvest an arterial segment at least 3 cm long and to examine a sufficient number of tissue sections. The histologic features of granulomatous inflammation are often quite subtle in temporal artery specimens. If the biopsy is declared negative by an experienced pathologist, the diagnosis of arteritic AION is highly unlikely and glucocorticoids should usually be discontinued.

Posterior Ischemic Optic Neuropathy

This is an uncommon cause of acute visual loss, induced by the combination of severe anemia and hypotension. Cases have been reported after major blood loss during surgery (especially in patients undergoing cardiac or lumbar spine operations), shock, gastrointestinal bleeding, and renal dialysis. The fundus usually appears normal, although optic disc swelling develops if the process extends anteriorly far enough to reach the globe. Vision can be salvaged in some patients by immediate blood transfusion and reversal of hypotension.

Optic Neuritis

This is a common inflammatory disease of the optic nerve. In the Optic Neuritis Treatment Trial (ONTT), the mean age of patients was 32 years, 77% were female, 92% had ocular pain (especially with eye movements), and 35% had optic disc swelling. In most patients, the demyelinating event was retrobulbar and the ocular fundus appeared normal on initial examination **(Fig. 32-10)**, although optic disc pallor slowly developed over subsequent months.

FIGURE 32-10

Retrobulbar optic neuritis is characterized by a normal fundus examination initially, hence the rubric "the doctor sees nothing, and the patient sees nothing." Optic atrophy develops after severe or repeated attacks. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Virtually all patients experience a gradual recovery of vision after a single episode of optic neuritis, even without treatment. This rule is so reliable that failure of vision to improve after a first attack of optic neuritis casts doubt on the original diagnosis. Treatment with high-dose IV methylprednisolone (250 mg every 6 h for 3 days) followed by oral prednisone (1 mg/kg per day for 11 days) makes no difference in ultimate acuity 6 months after the attack, but the recovery of visual function occurs more rapidly. Therefore, when visual loss is severe (worse than 20/100), IV followed by PO glucocorticoids are often recommended.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

neuritis. If two or more plaques are present on initial imaging, treatment should be considered to prevent the development of additional

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

but the recovery of visual function occurs more rapidly. Therefore, when visual loss is severe (worse than 20/100), IV followed by PO glucocorticoids are often recommended.

For some patients, optic neuritis remains an isolated event. However, the ONTT showed that the 15-year cumulative probability of developing clinically definite multiple sclerosis after optic neuritis is 50%. A brain magnetic resonance (MR) scan is advisable in every patient with a first attack of optic neuritis. If two or more plaques are present on initial imaging, treatment should be considered to prevent the development of additional demyelinating lesions **(Chap. 444)**.

A particularly severe optic neuritis, often involving a long segment of nerve, occurs in neuromyelitis optica (NMO); it may be bilateral and associated with myelitis. NMO can occur as a primary disorder, in the setting of systemic autoimmune disease, or rarely, as a paraneoplastic condition. Detection of circulating antibodies directed against aquaporin-4 or myelin oligodendrocyte glycoprotein (MOG) is diagnostic. Treatment for acute episodes consists of glucocorticoids followed by satralizumab, eculizumab, or inebilizumab to prevent relapse. **Neuromyelitis optica is discussed in detail in Chap. 445.**

LEBER'S HEREDITARY OPTIC NEUROPATHY

This disease usually affects young men, causing progressive, painless, severe central visual loss in one eye, followed weeks to years later by the same process in the other eye. Acutely, the optic disc appears mildly plethoric with surface capillary telangiectasias but no vascular leakage on fluorescein angiography. Eventually, optic atrophy ensues. Leber's optic neuropathy is caused by a point mutation at codon 11778 in the mitochondrial gene encoding nicotinamide adenine dinucleotide dehydrogenase (NADH) subunit 4. Additional mutations responsible for the disease have been identified, most in mitochondrial genes that encode proteins involved in electron transport. Mitochondrial mutations that cause Leber's neuropathy are maternally inherited by all children, but for unknown reasons, only 10% of cases occur in females. Clinical trials of gene therapy for this condition have been unsuccessful.

Toxic Optic Neuropathy

This can result in acute visual loss with bilateral optic disc swelling and cecocentral scotomas. Cases have been reported from exposure to ethambutol, methyl alcohol (moonshine), ethylene glycol (antifreeze), or carbon monoxide. In toxic optic neuropathy, visual loss also can develop gradually and produce optic atrophy **(Fig. 32-11)** without a phase of acute optic disc edema. Many agents have been implicated in toxic optic neuropathy, but evidence supporting the association is often weak. The following is a partial list of potential offending drugs or toxins: disulfiram, ethchlorvynol, chloramphenicol, amiodarone, monoclonal anti-CD3 antibody, ciprofloxacin, digitalis, streptomycin, lead, arsenic, thallium, D-penicillamine, isoniazid, emetine, and sulfonamides. Metallosis (chromium, cobalt, nickel) from hip implant failure is a rare cause of toxic optic neuropathy. Deficiency states induced by starvation, malabsorption, alcoholism, or gastric bypass can lead to insidious visual loss. Thiamine, vitamin B_{12}, and folate levels should be checked in any patient with unexplained bilateral central scotomas and optic pallor.

FIGURE 32-11

Optic atrophy is not a specific diagnosis but refers to the combination of optic disc pallor, arteriolar narrowing, and nerve fiber layer destruction produced by a host of eye diseases, especially optic neuropathies. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Papilledema

This connotes bilateral optic disc swelling from raised intracranial pressure **(Fig. 32-12)**. Headache is a common but not invariable accompaniment. All other forms of optic disc swelling (e.g., from optic neuritis or ischemic optic neuropathy) should be called "optic disc edema." This convention is arbitrary but serves to avoid confusion. Often it is difficult to differentiate papilledema from other forms of optic disc edema by fundus examination alone. Transient visual obscurations are a classic symptom of papilledema. They occur in only one eye or simultaneously in both eyes. They usually last seconds but can persist longer. Obscurations follow abrupt shifts in posture or happen spontaneously. When obscurations are prolonged or spontaneous, the papilledema is more threatening. Visual acuity is not affected by papilledema unless the papilledema is severe, long-standing, or accompanied by macular edema and hemorrhage. Visual field testing shows enlarged blind spots and peripheral constriction (Fig. 32-3F). With unremitting papilledema, peripheral visual field loss progresses in an insidious fashion while the optic nerve develops atrophy. In this setting, reduction of optic disc swelling is an ominous sign of a dying nerve rather than an encouraging indication of resolving papilledema.

FIGURE 32-12

Papilledema means optic disc edema from raised intracranial pressure. This young woman developed acute papilledema, with hemorrhages and cotton-wool spots, as a rare side effect of treatment with tetracycline for acne. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

search for a dural venous sinus thrombosis or an arteriovenous shunt. If neuroradiologic studies are negative, the subarachnoid opening pressure should be measured in the lateral decubitus position by lumbar puncture. Inaccurate pressure readings are a common pitfall. An elevated pressure,



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

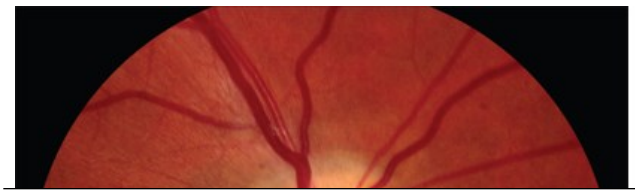
Evaluation of papilledema requires neuroimaging to exclude an intracranial lesion. Noninvasive MR vascular imaging may be useful in selected cases to search for a dural venous sinus thrombosis or an arteriovenous shunt. If neuroradiologic studies are negative, the subarachnoid opening pressure should be measured in the lateral decubitus position by lumbar puncture. Inaccurate pressure readings are a common pitfall. An elevated pressure, with normal cerebrospinal fluid, points by exclusion to the diagnosis of *pseudotumor cerebri* (idiopathic intracranial hypertension). Almost all patients are female, and most are obese. Treatment with a carbonic anhydrase inhibitor such as acetazolamide lowers intracranial pressure by reducing the production of cerebrospinal fluid and improves the visual fields. Weight reduction is vital: bariatric surgery should be considered in patients who cannot lose weight by diet control. If vision loss is severe or progressive, a shunt should be performed without delay to prevent blindness. Placement of a stent across the junction of the transverse and sigmoid dural sinuses, where stenosis is usually present, has emerged as a new treatment option. Optic nerve sheath fenestration is a less effective approach and does not address other neurologic symptoms. Occasionally, fulminant papilledema produces rapid onset of blindness. In such patients, emergency surgery should be performed to install a shunt.

Optic Disc Drusen

These are refractile, glittering particles within the substance of the optic nerve head **(Fig. 32-13)**. They are unrelated to drusen of the retina, which occur in age-related macular degeneration. Optic disc drusen are most common in people of northern European descent. Their diagnosis is obvious when they are visible on the surface of the optic disc. However, in many patients, they are hidden beneath the surface, producing pseudopapilledema. It is important to recognize optic disc drusen to avoid an unnecessary evaluation for papilledema. When optic disc drusen are buried, B-ultrasound is the most sensitive way to detect them. They appear hyperechoic because they contain calcium. They are also visible on computed tomography (CT) or optical coherence tomography (OCT), a technique for acquiring cross-section images of the retina. In most patients, optic disc drusen are an incidental, innocuous finding, but they can produce visual obscurations. On perimetry, they give rise to enlarged blind spots and arcuate scotomas from damage to the optic disc. With increasing age, drusen tend to become more exposed on the disc surface as optic atrophy develops. Hemorrhage, choroidal neovascular membrane, and AION are more likely to occur in patients with optic disc drusen. No treatment is available.

FIGURE 32-13

Optic disc drusen are calcified, mulberry-like deposits of unknown etiology within the optic disc, giving rise to "pseudopapilledema." *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Vitreous Degeneration

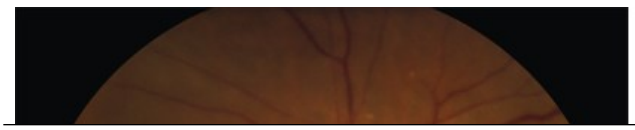
This occurs in all individuals with advancing age, leading to visual symptoms. Opacities develop in the vitreous, casting annoying shadows on the retina. As the eye moves, these distracting "floaters" move synchronously, with a slight lag caused by inertia of the vitreous gel. Vitreous traction on the retina causes mechanical stimulation, resulting in perception of flashing lights. This photopsia is brief and is confined to one eye, in contrast to the bilateral, prolonged scintillations of cortical migraine. Contraction of the vitreous can result in sudden separation from the retina, heralded by an alarming shower of floaters and photopsia. This process, known as *vitreous detachment*, is a common involutional event in the elderly. It is not harmful unless it damages the retina. A careful examination of the dilated fundus is important in any patient complaining of floaters or photopsia to search for peripheral tears or holes. If such a lesion is found, laser application can forestall a retinal detachment. Occasionally a tear ruptures a retinal blood vessel, causing vitreous hemorrhage and sudden loss of vision. On attempted ophthalmoscopy the fundus is hidden by a dark haze of blood. Ultrasound is required to examine the interior of the eye for a retinal tear or detachment. If the hemorrhage does not resolve spontaneously, the vitreous can be removed surgically. Vitreous hemorrhage also results from the fragile neovascular vessels that proliferate on the surface of the retina in diabetes, sickle cell anemia, and other ischemic ocular diseases.

Retinal Detachment

This produces symptoms of floaters, flashing lights, and a scotoma in the peripheral visual field corresponding to the detachment **(Fig. 32-14)**. If the detachment includes the fovea, there is an afferent pupil defect and the visual acuity is reduced. In most eyes, retinal detachment starts with a hole, flap, or tear in the peripheral retina (rhegmatogenous retinal detachment). Patients with peripheral retinal thinning (lattice degeneration) are particularly vulnerable to this process. Once a break has developed in the retina, liquefied vitreous is free to enter the subretinal space, separating the retina from the pigment epithelium. The combination of vitreous traction on the retinal surface and passage of fluid behind the retina leads inexorably to detachment. Patients with a history of myopia, trauma, or prior cataract extraction are at greatest risk for retinal detachment. The diagnosis is confirmed by ophthalmoscopic examination of the dilated eye.

FIGURE 32-14

Retinal detachment appears as an elevated sheet of retinal tissue with folds. In this patient, the fovea was spared, so acuity was normal, but an inferior detachment produced a superior scotoma. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Classic Migraine

(See also Chap. 430) This usually occurs with a visual aura lasting about 20 min. In a typical attack, a small central disturbance in the field of vision marches toward the periphery, leaving a transient scotoma in its wake. The expanding border of migraine scotoma has a scintillating, dancing, or zigzag edge, resembling the bastions of a fortified city, hence the term *fortification spectra*. Patients' descriptions of fortification spectra vary widely and can be confused with amaurosis fugax. Migraine patterns usually last longer and are perceived in both eyes, whereas amaurosis fugax is briefer and occurs in only one eye. Migraine phenomena also remain visible in the dark or with the eyes closed. Generally, they are confined to either the right or the left visual hemifield, but sometimes, both fields are involved simultaneously. Patients often have a long history of stereotypic attacks. After the visual symptoms recede, headache develops in most patients.

Transient Ischemic Attacks

Vertebrobasilar insufficiency may result in acute homonymous visual symptoms. Many patients mistakenly describe symptoms in the left or right eye when in fact the symptoms are occurring in the left or right hemifield of both eyes. Interruption of blood supply to the visual cortex causes a sudden fogging or graying of vision, occasionally with flashing lights or other positive phenomena that mimic migraine. Cortical ischemic attacks are briefer in duration than migraine, occur in older patients, and are not followed by headache. There may be associated signs of brainstem ischemia, such as diplopia, vertigo, numbness, weakness, and dysarthria.

Stroke

Stroke occurs when interruption of blood supply from the posterior cerebral artery to the visual cortex is prolonged. The only finding on examination is a homonymous visual field defect that stops abruptly at the vertical meridian. Occipital lobe stroke usually is due to thrombotic occlusion of the vertebrobasilar system, embolus, or dissection. Lobar hemorrhage, tumor, abscess, and arteriovenous malformation are other common causes of hemianopic cortical visual loss.

Factitious (Functional, Nonorganic) Visual Loss

This is claimed by hysterics or malingerers. The latter account for the vast majority, seeking sympathy, special treatment, or financial gain by feigning loss of sight. The diagnosis is suspected when the history is atypical, physical findings are lacking or contradictory, inconsistencies emerge on testing, and a secondary motive can be identified. In our litigious society, the fraudulent pursuit of recompense has spawned an epidemic of factitious visual loss.

CHRONIC VISUAL LOSS

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The formation of cataract occurs more rapidly in patients with a history of uveitis, diabetes mellitus, ocular trauma, or vitrectomy. Cataracts are acquired in a variety of genetic diseases, such as myotonic dystrophy, neurofibromatosis type 2, and galactosemia. Radiation therapy and

Cataract

Cataract is a clouding of the lens sufficient to reduce vision. Most cataracts develop slowly as a result of aging, leading to gradual impairment of vision. The formation of cataract occurs more rapidly in patients with a history of uveitis, diabetes mellitus, ocular trauma, or vitrectomy. Cataracts are acquired in a variety of genetic diseases, such as myotonic dystrophy, neurofibromatosis type 2, and galactosemia. Radiation therapy and glucocorticoid treatment can induce cataract as a side effect. The cataracts associated with radiation or glucocorticoids have a typical posterior subcapsular location. Cataract can be detected by noting an impaired red reflex when viewing light reflected from the fundus with an ophthalmoscope or by examining the dilated eye with the slit lamp.

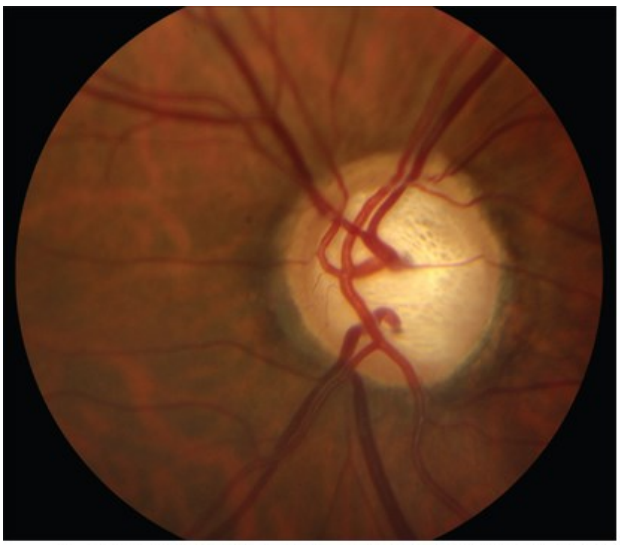
The only treatment for cataract is surgical extraction of the opacified lens. Millions of cataract operations are performed each year around the globe. The operation generally is done under local anesthesia on an outpatient basis. A plastic or silicone intraocular lens is placed within the empty lens capsule in the posterior chamber, substituting for the natural lens and leading to rapid recovery of sight. More than 95% of patients who undergo cataract extraction can expect an improvement in vision. In some patients, the lens capsule remaining in the eye after cataract extraction eventually turns cloudy, causing secondary loss of vision. A small opening, called a posterior capsulotomy, is made in the lens capsule with a laser to restore clarity.

Glaucoma

Glaucoma is a slowly progressive, insidious optic neuropathy that usually is associated with chronic elevation of intraocular pressure. After cataract, it is the most common cause of blindness in the world. It is especially prevalent in people of African descent. The mechanism by which raised intraocular pressure injures the optic nerve is not understood. Axons entering the inferotemporal and superotemporal aspects of the optic disc are damaged first, producing typical nerve fiber bundle defects called arcuate scotomas. As fibers are destroyed, the neural rim of the optic disc shrinks and the physiologic cup within the optic disc enlarges **(Fig. 32-15)**. This process is referred to as pathologic "cupping." The cup-to-disc diameter is expressed as a fraction (e.g., 0.2). The cup-to-disc ratio ranges widely in normal individuals, making it difficult to diagnose glaucoma reliably simply by observing an unusually large or deep optic cup. Careful documentation of serial examinations is helpful. In a patient with physiologic cupping, the large cup remains stable, whereas in a patient with glaucoma, it expands relentlessly over the years. Observation of progressive cupping and detection of an arcuate scotoma or a nasal step on computerized visual field testing is sufficient to establish the diagnosis of glaucoma. OCT reveals corresponding loss of fibers along the arcuate pathways in the nerve fiber layer.

FIGURE 32-15

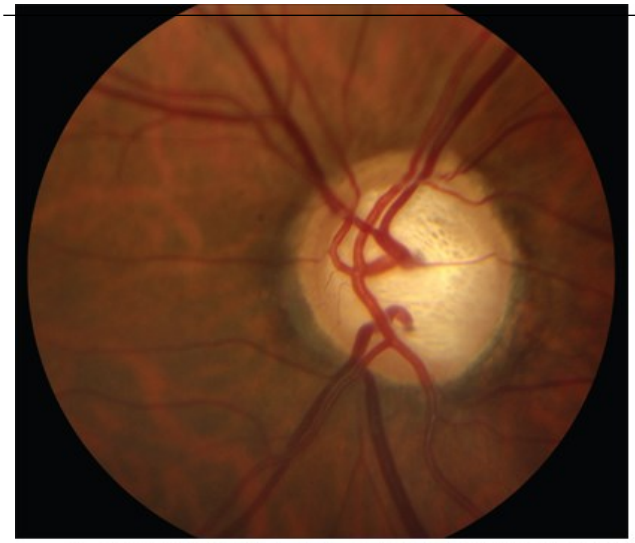
Glaucoma results in "cupping" as the neural rim is destroyed and the central cup becomes enlarged and excavated. The cup-to-disc ratio is about 0.8 in this patient. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser,

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

with open-angle glaucoma have an intraocular pressure within the normal range of 10–20 mmHg. For this so-called normal or low-tension form of glaucoma, high myopia is a risk factor.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The preponderance of patients with glaucoma have open anterior chamber angles. In most affected individuals, the intraocular pressure is elevated. The cause of elevated intraocular pressure is unknown, but it is associated with gene mutations in the heritable forms. Surprisingly, a third of patients with open-angle glaucoma have an intraocular pressure within the normal range of 10–20 mmHg. For this so-called normal or low-tension form of glaucoma, high myopia is a risk factor.

Chronic angle-closure glaucoma and chronic open-angle glaucoma are usually asymptomatic. Only acute angle-closure glaucoma causes a red or painful eye, from abrupt elevation of intraocular pressure. In all forms of glaucoma, foveal acuity is spared until end-stage disease is reached. For these reasons, severe and irreversible damage can occur before either the patient or the physician recognizes the diagnosis. Screening of patients for glaucoma by noting the cup-to-disc ratio on ophthalmoscopy and by measuring intraocular pressure is vital. Glaucoma is treated with topical adrenergic agonists, cholinergic agonists, beta blockers, prostaglandin analogues, and carbonic anhydrase inhibitors. Occasionally, systemic absorption of beta blocker from eyedrops can be sufficient to cause side effects of bradycardia, hypotension, heart block, bronchospasm, or depression. Laser treatment of the trabecular meshwork in the anterior chamber angle improves aqueous outflow from the eye. If medical or laser treatments fail to halt optic nerve damage from glaucoma, a filter must be constructed surgically (trabeculectomy) or a drainage device placed to release aqueous from the eye in a controlled fashion.

Macular Degeneration

This is a major cause of gradual, painless, bilateral central visual loss in the elderly. It occurs in a nonexudative (dry) form and an exudative (wet) form. Inflammation may be important in both forms of macular degeneration; susceptibility is associated with variants in the gene for complement factor H, an inhibitor of the alternative complement pathway. The nonexudative process begins with the accumulation of extracellular deposits called drusen underneath the retinal pigment epithelium. On ophthalmoscopy, they are pleomorphic but generally appear as small discrete yellow lesions clustered in the macula **(Fig. 32-16)**. With time, they become larger, more numerous, and confluent. The retinal pigment epithelium becomes focally detached and atrophic, causing visual loss by interfering with photoreceptor function. Treatment with vitamins C and E, beta-carotene, and zinc may retard dry macular degeneration.

FIGURE 32-16

Age-related macular degeneration consisting of scattered yellow drusen in the macula (dry form) and a crescent of fresh hemorrhage temporal to the fovea from a subretinal neovascular membrane (wet form). *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Exudative macular degeneration, which develops in only a minority of patients, occurs when neovascular vessels from the choroid grow through defects in Bruch's membrane and proliferate underneath the retinal pigment epithelium or the retina. Leakage from these vessels produces elevation of the retina, with distortion (metamorphopsia) and blurring of vision. Although the onset of these symptoms is usually gradual, bleeding from a subretinal choroidal neovascular membrane sometimes causes acute visual loss. Neovascular membranes can be difficult to see on fundus examination because they are located beneath the retina. Fluorescein angiography and OCT are extremely useful for their detection. Major or repeated hemorrhage under the retina from neovascular membranes results in fibrosis, development of a round (disciform) macular scar, and permanent loss of central vision.

A major therapeutic advance has occurred with the discovery that exudative macular degeneration can be treated with intraocular injection of antagonists to vascular endothelial growth factor. Bevacizumab, ranibizumab, aflibercept, or brolucizumab is administered by direct injection into the vitreous cavity, beginning on a monthly basis. These antibodies cause the regression of neovascular membranes by blocking the action of vascular endothelial growth factor, thereby improving visual acuity.

Central Serous Chorioretinopathy

This primarily affects males between the ages of 20 and 50 years. Leakage of serous fluid from the choroid causes small, localized detachment of the retinal pigment epithelium and the neurosensory retina. These detachments produce acute or chronic symptoms of metamorphopsia and blurred vision when the macula is involved. They are difficult to visualize with a direct ophthalmoscope because the detached retina is transparent and only slightly elevated. OCT shows fluid beneath the retina, and fluorescein angiography shows dye streaming into the subretinal space. The cause of central serous chorioretinopathy is unknown. Symptoms may resolve spontaneously if the retina reattaches, but recurrent detachment is common. Laser photocoagulation has benefited some patients with this condition.

Diabetic Retinopathy

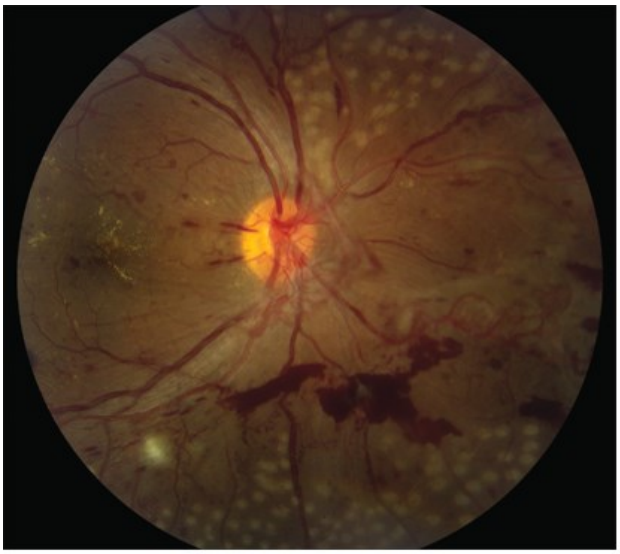
A rare disease until 1921, when the discovery of insulin resulted in a dramatic improvement in life expectancy for patients with diabetes mellitus, diabetic retinopathy is now a leading cause of blindness in the United States. The retinopathy takes years to develop but eventually appears in nearly all cases. Regular surveillance of the dilated fundus is crucial for any patient with diabetes. In advanced diabetic retinopathy, the proliferation of neovascular vessels leads to blindness from vitreous hemorrhage, retinal detachment, and glaucoma **(Fig. 32-17)**. These complications can be avoided in most patients by administration of panretinal laser photocoagulation at the appropriate point in the evolution of the disease. Anti-vascular endothelial growth factor antibody treatment is equally effective, but intraocular injections must be given repeatedly. **For further discussion of the manifestations and management of diabetic retinopathy, see Chaps. 403–405.**

FIGURE 32-17

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Copyright © McGraw Hill Education. All rights reserved.)

Proliferative diabetic retinopathy in a 25-year-old man with an 18-year history of diabetes, showing neovascular vessels emanating from the optic disc, retinal and vitreous hemorrhage, cotton-wool spots, and macular exudate. Round spots in the periphery represent recently applied panretinal photocoagulation. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



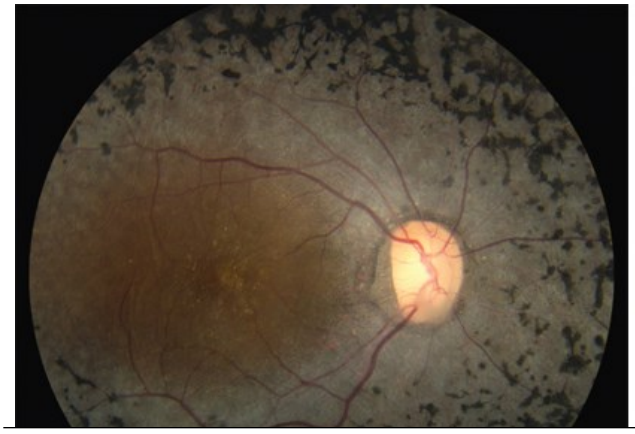
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Retinitis Pigmentosa

This is a general term for a disparate group of rod-cone dystrophies characterized by progressive night blindness, visual field constriction with a ring scotoma, loss of acuity, and an abnormal electroretinogram (ERG). It occurs sporadically or in an autosomal recessive, dominant, or X-linked pattern. Irregular black deposits of clumped pigment in the peripheral retina, called *bone spicules* because of their vague resemblance to the spicules of cancellous bone, give the disease its name **(Fig. 32-18)**. The name is actually a misnomer because retinitis pigmentosa is not an inflammatory process. Genetic testing usually identifies a mutation in the gene for rhodopsin, the rod photopigment, or in the gene for peripherin, a glycoprotein located in photoreceptor outer segments. Vitamin A (15,000 IU/d) slightly retards the deterioration of the ERG in patients with retinitis pigmentosa but has no beneficial effect on visual acuity or fields.

FIGURE 32-18

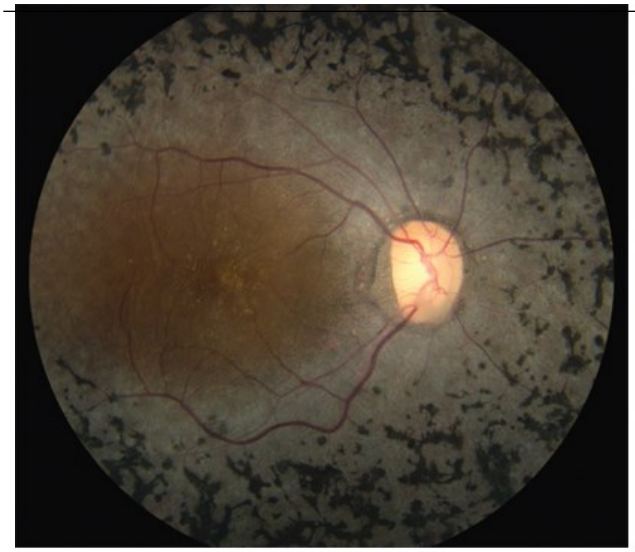
Retinitis pigmentosa with black clumps of pigment known as "bone spicules." The patient had peripheral visual field loss with sparing of central (macular) vision. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Leber's congenital amaurosis, a rare cone dystrophy, has been treated by replacement of the missing RPE65 protein through gene therapy, resulting in slight improvement in visual function. Some forms of retinitis pigmentosa occur in association with rare, hereditary systemic diseases (olivopontocerebellar degeneration, Bassen-Kornzweig disease, Kearns-Sayre syndrome, Refsum's disease). Chronic treatment with chloroquine, hydroxychloroquine, and phenothiazines (especially thioridazine) can produce visual loss from a toxic retinopathy that resembles retinitis pigmentosa. Patients receiving long-term treatment with hydroxychloroquine require regular eye examinations to monitor for potential development of a bull's eye maculopathy.

Epiretinal Membrane

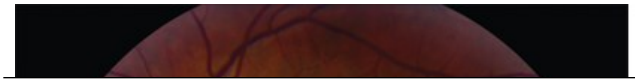
This is a fibrocellular tissue that grows across the inner surface of the retina, causing metamorphopsia and reduced visual acuity from distortion of the macula. A crinkled, cellophane-like membrane is visible on the retinal examination. Epiretinal membrane is most common in patients aged >50 years and is usually unilateral. Most cases are idiopathic, but some occur as a result of hypertensive retinopathy, diabetes, retinal detachment, or trauma. When visual acuity is reduced to the level of about 6/24 (20/80), vitrectomy and surgical peeling of the membrane to relieve macular puckering are recommended. Contraction of an epiretinal membrane sometimes gives rise to a *macular hole*. Most macular holes, however, are caused by local vitreous traction within the fovea. Vitrectomy can improve acuity in selected cases.

Melanoma and Other Tumors

Melanoma is the most common primary tumor of the eye **(Fig. 32-19)**. Approximately 2000 cases occur annually in the United States. It causes photopsia, an enlarging scotoma, and loss of vision. A small melanoma is often difficult to differentiate from a benign choroidal nevus. Serial examinations are required to document a malignant pattern of growth. Risk factors include light skin, hair, and eyes. Uveal origin accounts for 85% of cases. *GNAQ* and *GNA11* mutations are common. About half metastasize, mainly to the liver. Small and medium-sized tumors may be treated with radiation therapy; enucleation is the best treatment for large tumors. *Metastatic tumors* to the eye outnumber primary tumors. Breast and lung carcinomas have a special propensity to spread to the choroid or iris. Leukemia and lymphoma also commonly invade ocular tissues. Sometimes their only sign on eye examination is cellular debris in the vitreous, which can masquerade as a chronic posterior uveitis.

FIGURE 32-19

Melanoma of the choroid, appearing as an elevated dark mass in the inferior fundus, with overlying hemorrhage. The black line denotes the plane of the optical coherence tomography scan (*below*) showing the subretinal tumor. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

In a patient with vision loss, CT or MR scanning should be considered if the cause remains unknown after careful review of the history, visual fields, and thorough examination of the eye. Optic nerve sheath meningioma is a common retrobulbar tumor. It produces the classic triad of optociliary shunt vessels, optic atrophy, and progressive visual loss. Optic disc swelling and proptosis are also frequent signs. Optic nerve glioma in young patients is usually a pilocytic astrocytoma and has a good prognosis for preservation of vision, especially in neurofibromatosis type 1 **(Chap. 90)**. In adults, optic nerve glioma is rare and highly malignant. Chiasmal tumors (pituitary adenoma, meningioma, craniopharyngioma) produce visual loss with few objective findings except for optic disc pallor. Loss of the temporal visual field in each eye is typically described, but in fact, patients complain of vision loss in just one eye. A high degree of vigilance is necessary to avoid missing chiasmal tumors. Although symptoms progress gradually, in rare instances, the sudden expansion of a pituitary adenoma from infarction and bleeding (*pituitary apoplexy*) causes acute retrobulbar visual loss, with headache, nausea, and ocular motor nerve palsies.

PROPTOSIS

When the globes appear asymmetric, the clinician must first decide which eye is abnormal. Is one eye recessed within the orbit (*enophthalmos*), or is the other eye protuberant (*exophthalmos*, or *proptosis*)? A small globe or Horner's syndrome can give the appearance of enophthalmos. True enophthalmos occurs commonly after trauma, from atrophy of retrobulbar fat, or from fracture of the orbital floor. The position of the eyes within the orbits is measured by using a Hertel exophthalmometer, a handheld instrument that records the position of the anterior corneal surface relative to the lateral orbital rim. If this instrument is not available, relative eye position can be judged by bending the patient's head forward and looking down upon the orbits. A proptosis of only 2 mm in one eye is detectable from this perspective. The development of proptosis implies a space-occupying lesion in the orbit and usually warrants CT or MR imaging.

Graves' Ophthalmopathy

This is the leading cause of proptosis in adults **(Chap. 382)**. The proptosis is often asymmetric and can even appear to be unilateral. Orbital inflammation and engorgement of the extraocular muscles, particularly the medial rectus and the inferior rectus, account for the protrusion of the globe. Corneal exposure, lid retraction, lid lag on downgaze, conjunctival injection, restriction of gaze, diplopia, and visual loss from optic nerve compression are cardinal symptoms. Graves' eye disease is a clinical diagnosis, but laboratory testing can be useful. The serum level of thyroid-stimulating immunoglobulins is often elevated. Orbital imaging usually reveals enlarged extraocular eye muscles, but not always. Topical lubricants, taping the eyelids closed at night, and moisture chambers are helpful to limit exposure of ocular tissues. Graves' ophthalmopathy can be treated with oral prednisone (60 mg/d) for 1 month, followed by a taper over several months, but worsening of symptoms upon glucocorticoid withdrawal is common. Infusions of teprotumumab, an inhibitor of the insulin-like growth factor I receptor, reduce proptosis and diplopia. Radiation therapy is not

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Orbital Pseudotumor

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

oral prednisone (60 mg/d) for 1 month, followed by a taper over several months, but worsening of symptoms upon glucocorticoid withdrawal is common. Infusions of teprotumumab, an inhibitor of the insulin-like growth factor I receptor, reduce proptosis and diplopia. Radiation therapy is not effective. Orbital decompression should be performed for severe, symptomatic exophthalmos or if visual function is reduced by optic nerve compression. In patients with diplopia, prisms or eye muscle surgery can be used to restore ocular alignment in primary gaze.

Orbital Pseudotumor

This is an idiopathic, inflammatory orbital syndrome that is distinguished from Graves' ophthalmopathy by the prominent complaint of pain. Other symptoms include diplopia, ptosis, proptosis, and orbital congestion. Evaluation for sarcoidosis, granulomatosis with polyangiitis, and other types of orbital vasculitis or collagen-vascular disease is negative. Imaging often shows swollen eye muscles (orbital myositis) with enlarged tendons. By contrast, in Graves' ophthalmopathy, the tendons of the eye muscles usually are spared. The Tolosa-Hunt syndrome **(Chap. 441)** may be regarded as an extension of orbital pseudotumor through the superior orbital fissure into the cavernous sinus. The diagnosis of orbital pseudotumor is difficult. Biopsy of the orbit frequently yields nonspecific evidence of fat infiltration by lymphocytes, plasma cells, and eosinophils. A dramatic response to a therapeutic trial of systemic glucocorticoids indirectly provides the best confirmation of the diagnosis.

Orbital Cellulitis

This causes pain, lid erythema, proptosis, conjunctival chemosis, restricted motility, decreased acuity, afferent pupillary defect, fever, and leukocytosis. It often arises from the paranasal sinuses, especially by contiguous spread of infection from the ethmoid sinus through the lamina papyracea of the medial orbit. A history of recent upper respiratory tract infection, chronic sinusitis, thick mucus secretions, or dental disease is significant in any patient with suspected orbital cellulitis. Blood cultures should be obtained, but they are usually negative. Most patients respond to empirical therapy with broad-spectrum IV antibiotics. Occasionally, orbital cellulitis follows an overwhelming course, with massive proptosis, blindness, septic cavernous sinus thrombosis, and meningitis. To avert this disaster, orbital cellulitis should be managed aggressively in the early stages, with immediate imaging of the orbits and antibiotic therapy that includes coverage of methicillin-resistant *Staphylococcus aureus* (MRSA). Prompt surgical drainage of an orbital abscess or paranasal sinusitis is indicated if optic nerve function deteriorates despite antibiotics.

Tumors

Tumors of the orbit cause painless, progressive proptosis. The most common primary tumors are cavernous hemangioma, lymphangioma, neurofibroma, schwannoma, dermoid cyst, adenoid cystic carcinoma, optic nerve glioma, optic nerve meningioma, and benign mixed tumor of the lacrimal gland. Metastatic tumor to the orbit occurs frequently in breast carcinoma, lung carcinoma, and lymphoma. Diagnosis by fine-needle aspiration followed by urgent radiation therapy sometimes can preserve vision.

Carotid Cavernous Fistulas

With anterior drainage through the orbit, these fistulas produce proptosis, diplopia, glaucoma, and corkscrew, arterialized conjunctival vessels. Direct fistulas usually result from trauma. They are easily diagnosed because of the prominent signs produced by high-flow, high-pressure shunting. Indirect fistulas, or dural arteriovenous malformations, are more likely to occur spontaneously, especially in older women. The signs are more subtle, and the diagnosis frequently is missed. The combination of slight proptosis, diplopia, enlarged muscles, and an injected eye often is mistaken for thyroid ophthalmopathy. A bruit heard upon auscultation of the head or reported by the patient is a valuable diagnostic clue. Imaging shows an enlarged superior ophthalmic vein in the orbits. Carotid cavernous shunts can be eliminated by intravascular embolization.

PTOSIS

Blepharoptosis

This is an abnormal drooping of the eyelid. Unilateral or bilateral ptosis can be congenital, from dysgenesis of the levator palpebrae superioris, or from abnormal insertion of its aponeurosis into the eyelid. Acquired ptosis can develop so gradually that the patient is unaware of the problem. Inspection of old photographs is helpful in dating the onset. A history of prior trauma, eye surgery, contact lens use, diplopia, systemic symptoms (e.g., dysphagia or peripheral muscle weakness), or a family history of ptosis should be sought. Fluctuating ptosis that worsens late in the day is typical of myasthenia gravis. Ptosis evaluation should focus on evidence for proptosis, eyelid masses or deformities, inflammation, pupil inequality, or limitation of motility. The width of the palpebral fissures is measured in primary gaze to determine the degree of ptosis. The ptosis will be underestimated if the patient compensates by lifting the brow with the frontalis muscle.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

This occurs in many elderly patients from stretching and redundancy of eyelid skin and subcutaneous fat (dermatochalasis). The extra weight of these sagging tissues causes the lid to droop. Enlargement or deformation of the eyelid from infection, tumor, trauma, or inflammation also results in ptosis

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

of motility. The width of the palpebral fissures is measured in primary gaze to determine the degree of ptosis. The ptosis will be underestimated if the patient compensates by lifting the brow with the frontalis muscle.

Mechanical Ptosis

This occurs in many elderly patients from stretching and redundancy of eyelid skin and subcutaneous fat (dermatochalasis). The extra weight of these sagging tissues causes the lid to droop. Enlargement or deformation of the eyelid from infection, tumor, trauma, or inflammation also results in ptosis on a purely mechanical basis.

Aponeurotic Ptosis

This is an acquired dehiscence or stretching of the aponeurotic tendon, which connects the levator muscle to the tarsal plate of the eyelid. It occurs commonly in older patients, presumably from loss of connective tissue elasticity. Aponeurotic ptosis is also a common sequela of eyelid swelling from infection or blunt trauma to the orbit, cataract surgery, or contact lens use.

Myogenic Ptosis

The causes of *myogenic ptosis* include myasthenia gravis **(Chap. 448)** and a number of rare myopathies that manifest with ptosis. The term *chronic progressive external ophthalmoplegia* refers to a spectrum of systemic diseases caused by mutations of mitochondrial DNA. As the name implies, the most prominent findings are symmetric, slowly progressive ptosis and limitation of eye movements. In general, diplopia is a late symptom because all eye movements are reduced equally. In the *Kearns-Sayre* variant, retinal pigmentary changes and abnormalities of cardiac conduction develop. Peripheral muscle biopsy shows characteristic “ragged-red fibers.” *Oculopharyngeal dystrophy* is a distinct autosomal dominant disease with onset in middle age, characterized by ptosis, limited eye movements, and trouble swallowing. *Myotonic dystrophy*, another autosomal dominant disorder, causes ptosis, ophthalmoparesis, cataract, and pigmentary retinopathy. Patients have muscle wasting, myotonia, frontal balding, and cardiac abnormalities.

Neurogenic Ptosis

This results from a lesion affecting the innervation to either of the two muscles that open the eyelid: Müller’s muscle or the levator palpebrae superioris. Examination of the pupil helps distinguish between these two possibilities. In Horner’s syndrome, the eye with ptosis has a smaller pupil and the eye movements are full. In an oculomotor nerve palsy, the eye with the ptosis has a larger or a normal pupil. If the pupil is normal but there is limitation of adduction, elevation, and depression, a pupil-sparing oculomotor nerve palsy is likely (see next section). Rarely, a lesion affecting the small, central subnucleus of the oculomotor complex will cause bilateral ptosis with normal eye movements and pupils.

DOUBLE VISION (DIPLOPIA)

The first point to clarify is whether diplopia persists in either eye after the opposite eye is covered. If it does, the diagnosis is monocular diplopia. The cause is usually intrinsic to the eye and therefore has no dire implications for the patient. Corneal aberrations (e.g., keratoconus, pterygium), uncorrected refractive error, cataract, or foveal traction may give rise to monocular diplopia. Occasionally, it is a symptom of malingering or psychiatric disease. Diplopia alleviated by covering one eye is binocular diplopia and is caused by disruption of ocular alignment. Inquiry should be made into the nature of the double vision (purely side-by-side versus partial vertical displacement of images), mode of onset, duration, intermittency, diurnal variation, and associated neurologic or systemic symptoms. If the patient has diplopia while being examined, motility testing should reveal a deficiency corresponding to the patient’s symptoms. However, subtle limitation of ocular excursions is often difficult to detect. For example, a patient with a slight left abducens nerve paresis may appear to have full eye movements despite a complaint of horizontal diplopia upon looking to the left. In this situation, the cover test provides a more sensitive method for demonstrating the ocular misalignment. It should be conducted in primary gaze and then with the head turned and tilted in each direction while the patient fixates a central, distant target. In the above example, a cover test with the head turned to the right bringing the eyes into left gaze will maximize the fixation shift evoked by the cover test.

Occasionally, a cover test performed in an asymptomatic patient during a routine examination will reveal an ocular deviation. If the eye movements are full and the ocular misalignment is equal in all directions of gaze (comitant deviation), the diagnosis is strabismus. In this condition, which affects about 1% of the population, fusion is disrupted in infancy or early childhood. To avoid diplopia, retinal input from the nonfixating eye may be partially suppressed. In some children, this leads to impaired vision (amblyopia, or “lazy” eye) in the deviated eye.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

restriction is usually made by recognizing other associated signs and symptoms of local orbital disease. Dedicated, high-resolution orbital imaging is helpful when the cause of diplopia is not evident.

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

suppressed. In some children, this leads to impaired vision (amblyopia, or "lazy" eye) in the deviated eye.

Binocular diplopia results from a wide range of processes: infectious, neoplastic, metabolic, degenerative, inflammatory, and vascular. One must decide whether the diplopia is neurogenic in origin or is due to restriction of globe rotation by local disease in the orbit. Orbital pseudotumor, myositis, infection, tumor, thyroid disease, and muscle entrapment (e.g., from a blowout fracture) cause restrictive diplopia. The diagnosis of restriction is usually made by recognizing other associated signs and symptoms of local orbital disease. Dedicated, high-resolution orbital imaging is helpful when the cause of diplopia is not evident.

Myasthenia Gravis

(See also Chap. 448) This is a major cause of painless diplopia. The diplopia is often intermittent, variable, and not confined to any single ocular motor nerve distribution. The pupils are always normal. Serial observation of a fatigable ptosis, often accompanied by diplopia from fluctuating ocular misalignment, establishes the diagnosis. Many patients have a purely ocular form of the disease, with no evidence of systemic muscular weakness. Classically, the diagnosis was confirmed by an IV edrophonium injection, which produces a transient reversal of eyelid or eye muscle weakness, but this drug is discontinued in the United States. Blood tests for antibodies against the acetylcholine receptor or the MuSK protein are frequently negative in the purely ocular form of myasthenia gravis. *Botulism* from food or wound poisoning can mimic ocular myasthenia.

If restrictive orbital disease and myasthenia gravis are excluded, a lesion of a cranial nerve supplying innervation to the extraocular muscles is the most likely cause of binocular diplopia.

Oculomotor Nerve

The third cranial nerve innervates the medial, inferior, and superior recti; inferior oblique; levator palpebrae superioris; and the iris sphincter. Total palsy of the oculomotor nerve causes ptosis, a dilated pupil, and leaves the eye "down and out" because of the unopposed action of the lateral rectus and superior oblique. This combination of findings is obvious. More challenging is the diagnosis of early or partial oculomotor nerve palsy. In this setting, any combination of ptosis, pupil dilation, and weakness of the eye muscles supplied by the oculomotor nerve may be encountered. Frequent serial examinations during the rapidly evolving phase of the palsy help ensure that the diagnosis is not missed. The advent of an oculomotor nerve palsy with a pupil involvement, especially when accompanied by pain, suggests a compressive lesion, such as a tumor or circle of Willis aneurysm. Urgent neuroimaging should be obtained, along with a CT or MR angiogram. The resolution of these noninvasive techniques has advanced to the point that catheter angiography is rarely necessary to exclude an aneurysm.

A lesion of the oculomotor nucleus in the rostral midbrain produces signs that differ from those caused by a lesion of the nerve itself. There is bilateral ptosis because the levator muscle is innervated by a single central subnucleus. There is also weakness of the contralateral superior rectus, because it is supplied by the oculomotor nucleus on the other side. Occasionally both superior recti are weak. Isolated nuclear oculomotor palsy is rare. Usually, neurologic examination reveals additional signs that suggest brainstem damage from infarction, hemorrhage, tumor, or infection.

Injury to structures surrounding fascicles of the oculomotor nerve descending through the midbrain has given rise to a number of classic eponymic designations. In *Nothnagel's syndrome*, injury to the superior cerebellar peduncle causes ipsilateral oculomotor palsy and contralateral cerebellar ataxia. In *Benedikt's syndrome*, injury to the red nucleus results in ipsilateral oculomotor palsy and contralateral tremor, chorea, and athetosis. *Claude's syndrome* incorporates features of both of these syndromes, by injury to both the red nucleus and the superior cerebellar peduncle. Finally, in *Weber's syndrome*, injury to the cerebral peduncle causes ipsilateral oculomotor palsy with contralateral hemiparesis.

In the subarachnoid space, the oculomotor nerve is vulnerable to aneurysm, meningitis, tumor, infarction, and compression. In cerebral herniation, the nerve becomes trapped between the edge of the tentorium and the uncus of the temporal lobe. Oculomotor palsy also can result from midbrain torsion and hemorrhage during herniation. In the cavernous sinus, oculomotor palsy arises from carotid aneurysm, carotid cavernous fistula, cavernous sinus thrombosis, tumor (pituitary adenoma, meningioma, metastasis), herpes zoster infection, and the Tolosa-Hunt syndrome.

The etiology of an isolated, pupil-sparing oculomotor palsy often remains an enigma even after neuroimaging and extensive laboratory testing. Most cases are thought to result from microvascular infarction of the nerve somewhere along its course from the brainstem to the orbit. Usually, the patient complains of pain. Diabetes, hypertension, and vascular disease are major risk factors. Spontaneous recovery over a period of months is the rule. If this fails to occur or if new findings develop, the diagnosis of microvascular oculomotor nerve palsy should be reconsidered. Aberrant regeneration is common when the oculomotor nerve is injured by trauma or compression (tumor, aneurysm). Miswiring of sprouting fibers to the levator muscle and the rectus muscles results in elevation of the eyelid upon downgaze or adduction. The pupil also constricts upon attempted adduction, elevation, or depression of the globe. Aberrant regeneration is not seen after oculomotor palsy from microvascular infarct and hence vitiates that diagnosis.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The fourth cranial nerve originates in the midbrain, just caudal to the oculomotor nerve complex. Fibers exit the brainstem dorsally and cross to innervate the contralateral superior oblique. The principal actions of this muscle are to depress and intort the globe. A palsy therefore results in

depression of the globe. Aberrant regeneration is not seen after oculomotor palsy from microvascular infarct and hence vitiates that diagnosis.

Trochlear Nerve

The fourth cranial nerve originates in the midbrain, just caudal to the oculomotor nerve complex. Fibers exit the brainstem dorsally and cross to innervate the contralateral superior oblique. The principal actions of this muscle are to depress and intort the globe. A palsy therefore results in hypertropia and excyclotorsion. The cyclotorsion seldom is noticed by patients. Instead, they complain of vertical diplopia, especially upon reading or looking down. Vertical diplopia is exacerbated by tilting the head toward the side with the muscle palsy and alleviated by tilting it away. This "head tilt test" is a cardinal diagnostic feature. Review of old photographs will sometimes reveal a habitual head tilt, signifying a patient with a decompensated, congenital trochlear nerve palsy.

New, isolated trochlear nerve palsy results from all the causes listed above for the oculomotor nerve except aneurysm. The trochlear nerve is particularly apt to suffer injury after closed head trauma. The free edge of the tentorium impinges on the nerve during a concussive blow. Most isolated trochlear nerve palsies are idiopathic and hence are diagnosed by exclusion as "microvascular." Spontaneous improvement occurs over a period of months in most patients. A base-down prism (conveniently applied to the patient's glasses as a stick-on Fresnel lens) may serve as a temporary measure to alleviate diplopia. If the palsy does not resolve, the eyes can be realigned by weakening the inferior oblique muscle.

Abducens Nerve

The sixth cranial nerve innervates the lateral rectus muscle. A palsy produces horizontal diplopia, worse on gaze to the side of the lesion. A nuclear lesion has different consequences, because the abducens nucleus contains interneurons that project via the medial longitudinal fasciculus to the medial rectus subnucleus of the contralateral oculomotor complex. Therefore, an abducens nuclear lesion produces a complete lateral gaze palsy from weakness of both the ipsilateral lateral rectus and the contralateral medial rectus. *Foville's syndrome* after dorsal pontine injury includes lateral gaze palsy, ipsilateral facial palsy, and contralateral hemiparesis incurred by damage to descending corticospinal fibers. *Millard-Gubler syndrome* from ventral pontine injury is similar except for the eye findings. There is lateral rectus weakness only, instead of gaze palsy, because the abducens fascicle is injured rather than the nucleus. Infarct, tumor, hemorrhage, vascular malformation, and multiple sclerosis are the most common etiologies of brainstem abducens palsy.

After leaving the ventral pons, the abducens nerve runs forward along the clivus to pierce the dura at the petrous apex, where it enters the cavernous sinus. Along its subarachnoid course, it is susceptible to meningitis, tumor (meningioma, chordoma, carcinomatous meningitis), subarachnoid hemorrhage, trauma, and compression by aneurysm or dolichoectatic vessels. At the petrous apex, mastoiditis can produce deafness, pain, and ipsilateral abducens palsy (*Gradenigo's syndrome*). In the cavernous sinus, the nerve can be affected by carotid aneurysm, carotid cavernous fistula, tumor (pituitary adenoma, meningioma, nasopharyngeal carcinoma), herpes infection, and Tolosa-Hunt syndrome.

Unilateral or bilateral abducens palsy is a classic sign of raised intracranial pressure. The diagnosis can be confirmed if papilledema is observed on fundus examination. The mechanism is still debated but probably is related to rostral-caudal displacement of the brainstem. The same phenomenon accounts for abducens palsy from Chiari malformation or low intracranial pressure (e.g., after lumbar puncture, spinal anesthesia, or spontaneous dural cerebrospinal fluid leak).

Treatment of abducens palsy is aimed at prompt correction of the underlying cause. However, the cause remains obscure in many instances despite diligent evaluation. As was mentioned above for isolated trochlear or oculomotor palsy, most cases are assumed to represent microvascular infarcts because they often occur in the setting of diabetes or other vascular risk factors. Some cases may develop as a postinfectious mononeuritis (e.g., after a viral flu). Patching one eye, occluding one eyeglass lens with tape, or applying a temporary prism will provide relief of diplopia until the palsy resolves. If recovery is incomplete, eye muscle surgery nearly always can realign the eyes, at least in primary position. A patient with an abducens palsy that fails to improve should be reevaluated for an occult etiology (e.g., chordoma, carcinomatous meningitis, carotid cavernous fistula, myasthenia gravis). Skull base tumors are easily missed even on contrast-enhanced neuroimaging studies.

Multiple Ocular Motor Nerve Palsies

These should not be attributed to spontaneous microvascular events affecting more than one cranial nerve at a time. This remarkable coincidence does occur, especially in diabetic patients, but the diagnosis is made only in retrospect after all other diagnostic alternatives have been exhausted. Neuroimaging should focus on the cavernous sinus, superior orbital fissure, and orbital apex, where all three ocular motor nerves are in close proximity. In a diabetic or immunocompromised host, fungal infection (*Aspergillus*, Mucorales, *Cryptococcus*) is a common cause of multiple nerve

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Barré, produces ophthalmoplegia with areflexia and ataxia. Often the ataxia is mild, and the reflexes are normal. Antiganglioside antibodies (GQ1b) can be detected in about 50% of cases.

Neuroimaging should focus on the cavernous sinus, superior orbital fissure, and orbital apex, where all three ocular motor nerves are in close proximity. In a diabetic or immunocompromised host, fungal infection (*Aspergillus*, Mucorales, *Cryptococcus*) is a common cause of multiple nerve palsies. In a patient with systemic malignancy, carcinomatous meningitis is a likely diagnosis. Cytologic examination may be negative despite repeated sampling of the cerebrospinal fluid. The cancer-associated Lambert-Eaton myasthenic syndrome also can produce ophthalmoplegia. Giant cell (temporal) arteritis occasionally manifests as diplopia from ischemic palsies of extraocular muscles. Fisher's syndrome, an ocular variant of Guillain-Barré, produces ophthalmoplegia with areflexia and ataxia. Often the ataxia is mild, and the reflexes are normal. Antiganglioside antibodies (GQ1b) can be detected in about 50% of cases.

Supranuclear Disorders of Gaze

These are often mistaken for multiple ocular motor nerve palsies. For example, Wernicke's encephalopathy can produce nystagmus and a partial deficit of horizontal and vertical gaze that mimics a combined abducens and oculomotor nerve palsy. The disorder occurs in patients who are malnourished, alcoholic, or following bariatric surgery, and can be reversed by thiamine. Infarct, hemorrhage, tumor, multiple sclerosis, encephalitis, vasculitis, and Whipple's disease are other important causes of supranuclear gaze palsy. Disorders of vertical gaze, especially downward saccades, are an early feature of progressive supranuclear palsy. Smooth pursuit is affected later in the course of the disease. Parkinson's disease, Huntington's disease, and olivopontocerebellar degeneration also can affect vertical gaze.

The *frontal eye field* of the cerebral cortex is involved in generation of saccades to the contralateral side. After hemispheric stroke, the eyes usually deviate toward the lesioned side because of the unopposed action of the frontal eye field in the normal hemisphere. With time, this deficit resolves. Seizures generally have the opposite effect: the eyes deviate conjugately away from the irritative focus. *Parietal lesions* disrupt smooth pursuit of targets moving toward the side of the lesion. Bilateral parietal lesions produce *Bálint's syndrome*, which is characterized by impaired eye-hand coordination (optic ataxia), difficulty initiating voluntary eye movements (ocular apraxia), and visuospatial disorientation (simultanagnosia).

Horizontal Gaze

Descending cortical inputs mediating horizontal gaze ultimately converge at the level of the pons. Neurons in the paramedian pontine reticular formation are responsible for controlling conjugate gaze toward the same side. They project directly to the ipsilateral abducens nucleus. A lesion of either the paramedian pontine reticular formation or the abducens nucleus causes an ipsilateral conjugate gaze palsy. Lesions at either locus produce nearly identical clinical syndromes, with the following exception: vestibular stimulation (oculocephalic maneuver or caloric irrigation) will succeed in driving the eyes conjugately to the side in a patient with a lesion of the paramedian pontine reticular formation but not in a patient with a lesion of the abducens nucleus.

INTERNUCLEAR OPHTHALMOPLEGIA

This results from damage to the medial longitudinal fasciculus ascending from the abducens nucleus in the pons to the oculomotor nucleus in the midbrain (hence, "internuclear"). Damage to fibers carrying the conjugate signal from abducens interneurons to the contralateral medial rectus motoneurons results in a failure of adduction on attempted lateral gaze. For example, a patient with a left internuclear ophthalmoplegia (INO) will have slowed or absent adducting movements of the left eye **(Fig. 32-20)**. A patient with bilateral injury to the medial longitudinal fasciculus will have bilateral INO. Multiple sclerosis is the most common cause, although tumor, stroke, trauma, or any brainstem process may be responsible. *One-and-a-half syndrome* is due to a lesion of the medial longitudinal fasciculus combined with a lesion of either the abducens nucleus or the paramedian pontine reticular formation on the same side. The patient's only horizontal eye movement is abduction of the eye on the other side.

FIGURE 32-20

Left internuclear ophthalmoplegia (INO). ***A.*** In primary position of gaze, the eyes appear normal. ***B.*** Horizontal gaze to the left is intact. ***C.*** On attempted horizontal gaze to the right, the left eye fails to adduct. In mildly affected patients, the eye may adduct partially or more slowly than normal. Nystagmus is usually present in the abducted eye. ***D.*** T2-weighted axial magnetic resonance image through the pons showing a demyelinating plaque in the left medial longitudinal fasciculus (*arrow*). *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS



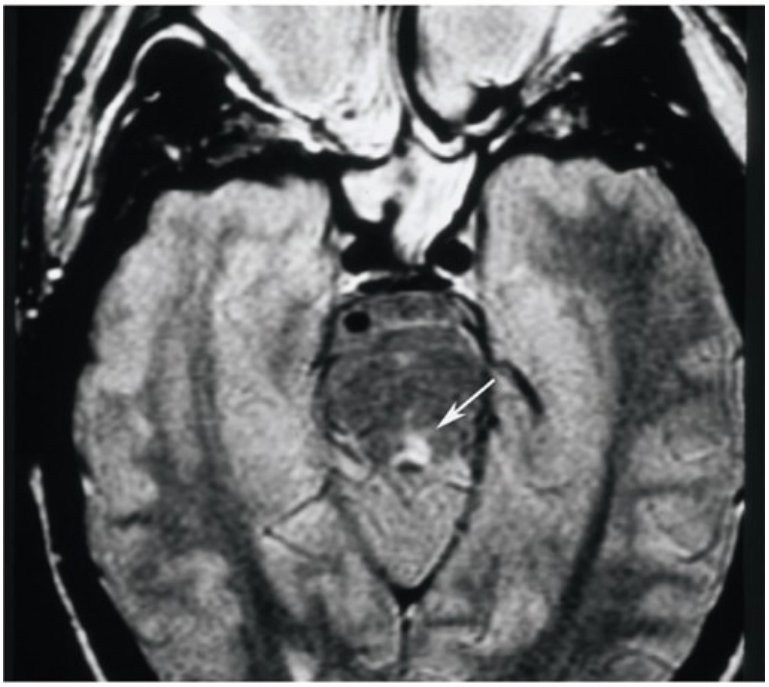
A



B



C



D

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser,

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

This is controlled at the level of the midbrain. The neuronal circuits affected in disorders of vertical gaze are not fully elucidated, but lesions of the

Access Provided by:
SILVERCHAIR

D

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Vertical Gaze

This is controlled at the level of the midbrain. The neuronal circuits affected in disorders of vertical gaze are not fully elucidated, but lesions of the rostral interstitial nucleus of the medial longitudinal fasciculus and the interstitial nucleus of Cajal cause supranuclear paresis of upgaze, downgaze, or all vertical eye movements. Distal basilar artery ischemia is the most common etiology. *Skew deviation* refers to a vertical misalignment of the eyes, usually constant in all positions of gaze. The finding has poor localizing value because skew deviation has been reported after lesions in widespread regions of the brainstem and cerebellum.

PARINAUD'S SYNDROME

Also known as dorsal midbrain syndrome, this is a distinct supranuclear vertical gaze disorder caused by damage to the posterior commissure. It is a classic sign of hydrocephalus from aqueductal stenosis. Pineal region or midbrain tumors, cysticercosis, and stroke also cause Parinaud's syndrome. Features include loss of upgaze (and sometimes downgaze), convergence-retraction nystagmus on attempted upgaze, downward ocular deviation ("setting sun" sign), lid retraction (Collier's sign), skew deviation, pseudoabducens palsy, and light-near dissociation of the pupils.

Nystagmus

This is a rhythmic oscillation of the eyes, occurring physiologically from vestibular and optokinetic stimulation or pathologically in a wide variety of diseases (Chap. 22). Abnormalities of the eyes or optic nerves, present at birth or acquired in childhood, can produce a complex, searching nystagmus with irregular pendular (sinusoidal) and jerk features. Examples are albinism, Leber's congenital amaurosis, and bilateral cataract. This nystagmus is commonly referred to as *congenital sensory nystagmus*. This is a poor term because even in children with congenital lesions, the nystagmus does not appear until weeks after birth. *Congenital motor nystagmus*, which looks similar to congenital sensory nystagmus, develops in the absence of any abnormality of the sensory visual system. Visual acuity also is reduced in congenital motor nystagmus, probably by the nystagmus itself, but seldom below a level of 20/200.

JERK NYSTAGMUS

This is characterized by a slow drift off the target, followed by a fast corrective saccade. By convention, the nystagmus is named after the quick phase. Jerk nystagmus can be downbeat, upbeat, horizontal (left or right), and torsional. The pattern of nystagmus may vary with gaze position. Some patients will be oblivious to their nystagmus. Others will complain of blurred vision or a subjective to-and-fro movement of the environment (oscillopsia) corresponding to the nystagmus. Fine nystagmus may be difficult to see on gross examination of the eyes. Observation of nystagmoid movements of the optic disc on ophthalmoscopy is a sensitive way to detect subtle nystagmus.

GAZE-EVOKED NYSTAGMUS

This is the most common form of jerk nystagmus. When the eyes are held eccentrically in the orbits, they have a natural tendency to drift back to primary position. The subject compensates by making a corrective saccade to maintain the deviated eye position. Many normal patients have mild gaze-evoked nystagmus. Exaggerated gaze-evoked nystagmus can be induced by drugs (sedatives, anticonvulsants, alcohol); muscle paresis; myasthenia gravis; demyelinating disease; and cerebellopontine angle, brainstem, and cerebellar lesions.

VESTIBULAR NYSTAGMUS

Vestibular nystagmus results from dysfunction of the labyrinth (Ménière's disease), vestibular nerve, or vestibular nucleus in the brainstem. Peripheral vestibular nystagmus often occurs in discrete attacks, with symptoms of nausea and vertigo. There may be associated tinnitus and hearing loss. Sudden shifts in head position may provoke or exacerbate symptoms.

DOWNBEAT NYSTAGMUS

Downbeat nystagmus results from lesions near the craniocervical junction (Chiari malformation, basilar invagination). It also has been reported in brainstem or cerebellar stroke, lithium or anticonvulsant intoxication, alcoholism, and multiple sclerosis. *Upbeat nystagmus* is associated with damage to the pontine tegmentum from stroke, demyelination, or tumor.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

This rare, dramatic disorder of eye movements consists of bursts of consecutive saccades (saccadomania). When the saccades are confined to the horizontal plane, the term *ocular flutter* is preferred. It can result from viral encephalitis, trauma, or a paraneoplastic effect of neuroblastoma, breast

brainstem or cerebellar stroke, lithium or anticonvulsant intoxication, alcoholism, and multiple sclerosis. *Upbeat nystagmus* is associated with damage to the pontine tegmentum from stroke, demyelination, or tumor.

Opsoclonus

This rare, dramatic disorder of eye movements consists of bursts of consecutive saccades (saccadomania). When the saccades are confined to the horizontal plane, the term *ocular flutter* is preferred. It can result from viral encephalitis, trauma, or a paraneoplastic effect of neuroblastoma, breast carcinoma, and other malignancies. It has also been reported as a benign, transient phenomenon in otherwise healthy patients.

FURTHER READING

Adamis AP et al: Building on the success of anti-vascular endothelial growth factor therapy: A vision for the next decade. Eye 34:1966, 2020. [PubMed: 32541890]

Douglas RS: Teprotumumab for the treatment of active thyroid eye disease. N Engl J Med 382:341, 2020. [PubMed: 31971679]

Dowling JE: Restoring vision to the blind. Science 368:827, 2020. [PubMed: 32439781]

Gross JG et al: Panretinal photocoagulation vs intravitreous ranibizumab for proliferative diabetic retinopathy. JAMA 314:2137, 2015. [PubMed: 26565927]

Jaffe GJ et al: Adalimumbab in patients with active noninfectious uveitis. N Engl J Med 375:932, 2016. [PubMed: 27602665]

Maeder ML: Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. Nat Med 25:229, 2019. [PubMed: 30664785]

Pioro MH: Primary care vasculitis: Polymyalgia rheumatica and giant cell arteritis. Prim Care 45:305, 2018. [PubMed: 29759126]

Stone JH et al: Trial of tocilizumab in giant-cell arteritis. N Engl J Med 377:317, 2017. [PubMed: 28745999]

Yanoff M, Duker J: *Ophthalmology*, 5th ed. Atlanta, Saunders, 2019.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 32: Disorders of the Eye, Jonathan C. Horton
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 34: Disorders of Hearing

Anil K. Lalwani

INTRODUCTION

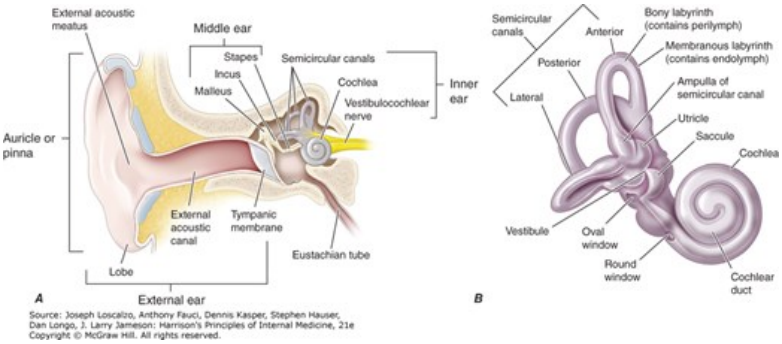
Hearing loss can present at any age and is one of the most common sensory disorders in humans. Nearly 10% of the adult population has some hearing loss, and one-third of individuals age >65 years have a hearing loss of sufficient magnitude to require a hearing aid.

PHYSIOLOGY OF HEARING

The function of the external and middle ear is to amplify sound to facilitate conversion of the mechanical energy of the sound wave into an electrical signal by the inner-ear hair cells, a process called mechanotransduction **(Fig. 34-1)**. Sound waves enter the external auditory canal and set the tympanic membrane (eardrum) in motion, which in turn moves the malleus, incus, and stapes of the middle ear. Movement of the footplate of the stapes causes pressure changes in the fluid-filled inner ear, eliciting a traveling wave in the basilar membrane of the cochlea. The tympanic membrane and the ossicular chain in the middle ear serve as an impedance-matching mechanism, improving the efficiency of energy transfer from air to the fluid-filled inner ear. In its absence, nearly 99.9% of the acoustical energy would be reflected and thus not heard. Instead, the eardrum and the ossicles boost the sound energy nearly 200-fold by the time it reaches the inner ear.

FIGURE 34-1

Ear anatomy. ***A.*** Drawing of modified coronal section through external ear and temporal bone, with structures of the middle and inner ear demonstrated. ***B.*** High-resolution view of inner ear.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Within the cochlea of the inner ear, there are two types of hair cells that aid in hearing: inner and outer. The inner and outer hair cells of the organ of Corti have different innervation patterns, but both are mechanoreceptors; they detect the mechanical energy of the acoustic signal and aid its conversion to an electrical signal that travels by the auditory nerve. The afferent innervation relates principally to the inner hair cells while the efferent innervation relates principally to the outer hair cells. The outer hair cells outnumber the inner hair cells by nearly 6:1 (20,000 vs 3500). The motility of the outer hair cells alters the micromechanics of the inner hair cells, creating a cochlear amplifier, which explains the exquisite sensitivity and frequency selectivity of the cochlea.

Stereocilia of the hair cells of the organ of Corti, which rests on the basilar membrane, are in contact with the tectorial membrane and are deformed by the traveling wave. The deformation stretches tiny filamentous connections (tip links) between stereocilia, leading to opening of ion channels, influx of potassium, and hair cell depolarization and subsequent neurotransmission. A point of maximal displacement of the basilar membrane is determined by the frequency of the stimulating tone. High-frequency tones cause maximal displacement of the basilar membrane near the base of the cochlea, whereas for low-frequency sounds, the point of maximal displacement is toward the apex of the cochlea.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

neurons alternate in response to particular phases of the cycle of the sound wave. Intensity is encoded by the amount of neural activity in individual neurons, the number of neurons that are active, and the specific neurons that are activated.

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Beginning in the cochlea, the frequency specificity is maintained at each point of the central auditory pathway: dorsal and ventral cochlear nuclei, trapezoid body, superior olivary complex, lateral lemniscus, inferior colliculus, medial geniculate body, and auditory cortex. At low frequencies, individual auditory nerve fibers can respond more or less synchronously with the stimulating tone. At higher frequencies, phase-locking occurs so that neurons alternate in response to particular phases of the cycle of the sound wave. Intensity is encoded by the amount of neural activity in individual neurons, the number of neurons that are active, and the specific neurons that are activated.

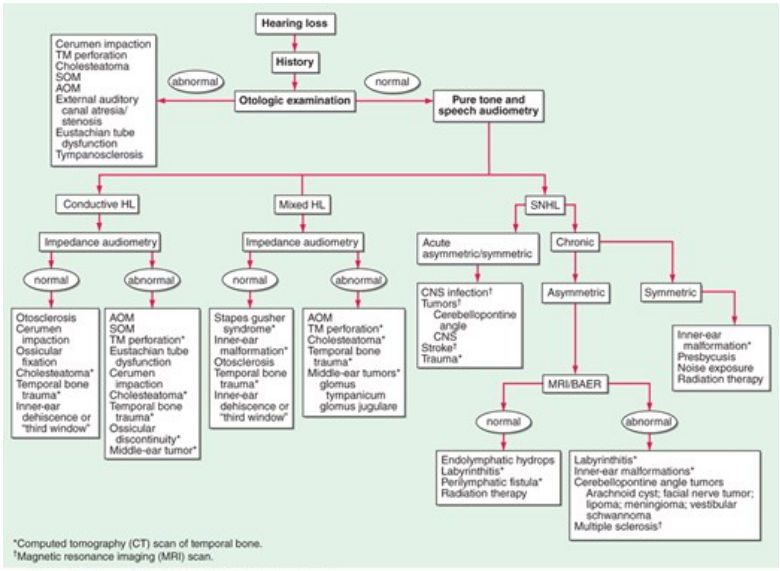
There is evidence that the right and left ears as well as the central nervous system may process speech asymmetrically. Generally, a sound is processed symmetrically from the peripheral to the central auditory system. However, a "right ear advantage" exists for dichotic listening tasks, in which subjects are asked to report on competing sounds presented to each ear. In most individuals, a perceptual right ear advantage for consonant-vowel syllables, stop consonants, and words also exists. Similarly, whereas central auditory processing for sounds is symmetric with minimal lateral specialization for the most part, speech processing is lateralized. There is specialization of the left auditory cortex for speech recognition and production, and of the right hemisphere for emotional and tonal aspects of speech. Left hemisphere dominance for speech is found in 95–98% of right-handed persons and 70–80% of left-handed persons.

DISORDERS OF THE SENSE OF HEARING

Hearing loss can result from disorders of the auricle, external auditory canal, middle ear, inner ear, or central auditory pathways **(Fig. 34-2)**. In general, lesions in the auricle, external auditory canal, or middle ear that impede the transmission of sound from the external environment to the inner ear cause conductive hearing loss, whereas lesions that impair mechanotransduction in the inner ear or transmission of the electrical signal along the eighth nerve to the brain cause sensorineural hearing loss.

FIGURE 34-2

An algorithm for the approach to hearing loss. AOM, acute otitis media; BAER, brainstem auditory-evoked response; CNS, central nervous system; HL, hearing loss; SNHL, sensorineural hearing loss; SOM, serous otitis media; TM, tympanic membrane.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Conductive Hearing Loss

The external ear, the external auditory canal, and the middle-ear apparatus are designed to collect and amplify sound and efficiently transfer the mechanical energy of the sound wave to the fluid-filled cochlea. Factors that obstruct the transmission of sound or dampen the acoustic energy result in conductive hearing loss. Conductive hearing loss can occur from obstruction of the external auditory canal by cerumen, debris, and foreign bodies; swelling of the lining of the canal; atresia or neoplasms of the canal; perforations of the tympanic membrane; disruption of the ossicular chain, as occurs with necrosis of the long process of the incus in trauma or infection; otosclerosis; or fluid, scarring, or neoplasms in the middle ear. Rarely, inner-ear malformations or pathologies that create a "third window" in the inner ear such as superior semicircular canal dehiscence, lateral

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Eustachian tube dysfunction is extremely common in adults and may predispose to acute otitis media (AOM) or serous otitis media (SOM). Recently,

ACCESS Medicine®
Access Provided by: SILVERCHAIR

occurs with necrosis of the long process of the incus in trauma or infection, otosclerosis, or fluid, scarring, or neoplasms in the middle ear. Rarely, inner-ear malformations or pathologies that create a "third window" in the inner ear such as superior semicircular canal dehiscence, lateral semicircular canal dysplasia, incomplete partition of the inner ear, and large vestibular aqueduct, are also associated with conductive hearing loss. This pathologic third window is associated with loss of mechanical energy associated with the sound wave leading to conductive hearing loss (see below).

Eustachian tube dysfunction is extremely common in adults and may predispose to acute otitis media (AOM) or serous otitis media (SOM). Recently, Eustachian tube balloon dilation has been shown to relieve acquired inflammatory obstruction of the Eustachian tube orifice and improve symptoms due to Eustachian tube dysfunction. Trauma, AOM, and chronic otitis media are the usual factors responsible for tympanic membrane perforation. While small perforations often heal spontaneously, larger defects usually require surgical intervention. Tympanoplasty is highly effective (>90%) in the repair of tympanic membrane perforations. Otoscopy is usually sufficient to diagnose AOM, SOM, chronic otitis media, cerumen impaction, tympanic membrane perforation, and Eustachian tube dysfunction; tympanometry and Eustachian tube function testing can be useful to confirm the clinical suspicion of these conditions.

Cholesteatoma, a benign tumor composed of stratified squamous epithelium in the middle ear or mastoid, occurs frequently in adults, often in the setting of severe Eustachian tube dysfunction. This is a slowly growing lesion that destroys bone and normal ear tissue. Theories of pathogenesis include traumatic immigration and invasion of squamous epithelium through a retraction pocket of the tympanic membrane, implantation of squamous epithelia in the middle ear through a perforation or surgery, and metaplasia following chronic infection and irritation. A chronically draining ear that fails to respond to appropriate antibiotic therapy should raise suspicion of a cholesteatoma. On examination, there is often a perforation of the tympanic membrane filled with cheesy white squamous debris. The presence of an aural polyp obscuring the tympanic membrane is highly suggestive of an underlying cholesteatoma. Conductive hearing loss secondary to ossicular erosion is common. Bony destruction visualized on CT of the temporal bone is also highly suggestive of cholesteatoma. Surgery is required to remove this destructive process and reconstruct the ossicles.

Conductive hearing loss with a normal ear canal and intact tympanic membrane suggests either ossicular pathology or the presence of a "third window" in the inner ear (see below). Fixation of the stapes from *otosclerosis* is a common cause of low-frequency conductive hearing loss. It occurs equally in men and women and is inherited as an autosomal dominant trait with incomplete penetrance; in some cases, it may be a manifestation of osteogenesis imperfecta. Hearing impairment usually presents between the late teens and the forties. In women, the otosclerotic process is accelerated during pregnancy, and the hearing loss is often first noticeable at this time. A hearing aid or a simple outpatient surgical procedure (stapedectomy) can provide excellent auditory rehabilitation. Extension of otosclerosis beyond the stapes footplate to involve the cochlea (cochlear otosclerosis) can lead to mixed or sensorineural hearing loss. Fluoride therapy to prevent hearing loss from cochlear otosclerosis is of uncertain value.

Disorders that lead to the formation of a pathologic "third window" in the inner ear can be associated with conductive hearing loss. There are normally two major openings, or windows, that connect the inner ear with the middle ear and serve as conduits for transmission of sound; these are, respectively, the oval and round windows. A third window is formed where the normally hard otic bone surrounding the inner ear is eroded; dissipation of the acoustic energy at the third window is responsible for the "inner-ear conductive hearing loss." The superior semicircular canal dehiscence syndrome resulting from erosion of the otic bone over the superior circular canal can present with conductive hearing loss that mimics otosclerosis. A common symptom is vertigo evoked by loud sounds (Tullio phenomenon), by Valsalva maneuvers that change middle-ear pressure, or by applying positive pressure on the tragus (the cartilage anterior to the external opening of the ear canal). Patients with this syndrome also complain of fullness of the ear, pulsatile tinnitus, and being able to hear the movement of their eyes and neck. A large jugular bulb or jugular bulb diverticulum can create a "third window" by eroding into the vestibular aqueduct or posterior semicircular canal; the symptoms are similar to those of the superior semicircular canal dehiscence syndrome. Other inner-ear malformations such as lateral semicircular canal dysplasia, large vestibular aqueduct, or incomplete partition seen in stapes gusher syndrome can also be associated with inner-ear conductive hearing loss as a result of the third window. Low activation threshold on the vestibular-evoked myogenic potential test (VEMP test, see below) and inner-ear erosion on CT are diagnostic. Recalcitrant vertigo and dizziness may respond to surgical repair of the dehiscence.

Sensorineural Hearing Loss

Sensorineural hearing loss results from either damage to the mechanotransduction apparatus of the cochlea or disruption of the electrical conduction pathway from the inner ear to the brain. Thus, injury to hair cells, supporting cells, auditory neurons, or the central auditory pathway can cause sensorineural hearing loss. Damage to the hair cells of the organ of Corti may be caused by intense noise, viral infections, ototoxic drugs (e.g., salicylates, quinine and its synthetic analogues, aminoglycoside antibiotics, loop diuretics such as furosemide and ethacrynic acid, and cancer chemotherapeutic agents such as cisplatin), fractures of the temporal bone, meningitis, cochlear otosclerosis (see above), Ménière's disease, and

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Noise-Induced Hearing Loss

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

salicylates, quinine and its synthetic analogues, aminoglycoside antibiotics, loop diuretics such as furosemide and ethacrynic acid, and cancer chemotherapeutic agents such as cisplatin), fractures of the temporal bone, meningitis, cochlear otosclerosis (see above), Ménière's disease, and aging. Congenital malformations of the inner ear may be the cause of hearing loss in some adults. Genetic predisposition alone or in concert with environmental exposures may also be responsible (see below).

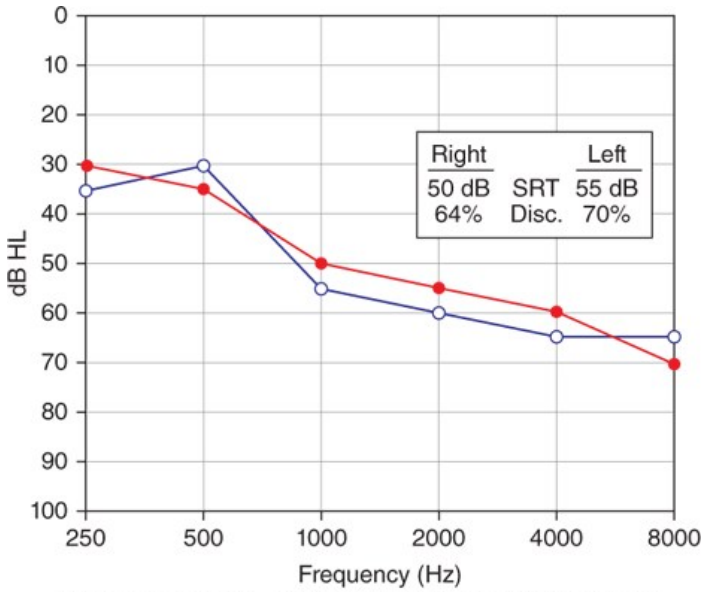
Noise-Induced Hearing Loss

Exposure to loud noise, either a short burst or over a more prolonged period of time, can lead to noise-induced hearing loss. Acute exposure to noise can lead to either temporary or permanent threshold shifts, depending on the intensity and duration of sound, due to hair cell injury and/or death. Typically, with permanent hearing loss there is a "noise notch" with elevated hearing thresholds at 3000–4000 Hz. More recently, loud noise exposure has also been associated with "hidden hearing loss"—hidden, because routine audiometry shows the pure tone hearing to be normal. Patients usually complain of not being able to hear clearly and are more bothered by the presence of background noise. In contrast to hair cell loss, hidden hearing loss is thought to be due to loss of auditory synapses on hair cells following noise exposure. In an increasingly noisy world, avoiding acoustic trauma with earplugs or earmuffs is highly recommended to prevent noise-induced or hidden hearing loss.

Presbycusis (age-associated hearing loss) is the most common cause of sensorineural hearing loss in adults. It is estimated to affect over half of adults aged >75 years in the United States, a population that is expected to double in size over the next 40 years. In the early stages, it is characterized by symmetric, gentle to sharply sloping, high-frequency hearing loss **(Fig. 34-3)**. With progression, the hearing loss involves all frequencies. More importantly, the hearing impairment is associated with significant loss in clarity. There is a loss of discrimination for phonemes, recruitment (abnormal growth of loudness), and particular difficulty in understanding speech in noisy environments such as at restaurants and social events. Poor hearing is also associated with an increased incidence of cognitive impairment, rate of cognitive decline, and falls. In the elderly, left untreated, hearing loss leads to diminished quality of life, and has been shown to increase overall morbidity and mortality through falls and accidents. Hearing aids are helpful in enhancing the signal-to-noise ratio by amplifying sounds that are close to the listener. Hearing aid use has been shown to reduce cognitive decline and risk of falls. Although hearing aids are able to amplify sounds, they cannot restore the clarity of hearing. Thus, amplification with hearing aids may provide only limited rehabilitation once the word recognition score deteriorates below 50%. Cochlear implants are the treatment of choice when hearing aids prove inadequate, even when hearing loss is incomplete (see below).

FIGURE 34-3

Presbycusis or age-related hearing loss. The audiogram shows a moderate to severe downsloping sensorineural hearing loss typical of presbycusis. The loss of high-frequency hearing is associated with a decreased speech discrimination score; consequently, patients complain of lack of clarity of hearing, especially in a noisy background. HL, hearing threshold level; SRT, speech reception threshold.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Ménière's disease is characterized by episodic vertigo, fluctuating sensorineural hearing loss, tinnitus, and aural fullness. An absence of vertigo is

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

frequently in the fifth decade of life but may also occur in young adults or the elderly. Histologically, there is distention of the endolymphatic system (endolymphatic hydrops) leading to degeneration of vestibular and cochlear hair cells. This may result from endolymphatic sac dysfunction secondary

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS
Copyright © McGraw Hill. All rights reserved.

Ménière's disease is characterized by episodic vertigo, fluctuating sensorineural hearing loss, tinnitus, and aural fullness. An absence of vertigo is inconsistent with the diagnosis of Ménière's disease, and the presence of fluctuating sensorineural hearing loss, tinnitus, and fullness without vertigo is more suggestive of cochlear hydrops. Tinnitus and/or deafness may be absent during the initial attacks of vertigo, but invariably appear as the disease progresses and increases in severity during acute attacks. The annual incidence of Ménière's disease is 0.5–7.5 per 1000; onset is most frequently in the fifth decade of life but may also occur in young adults or the elderly. Histologically, there is distention of the endolymphatic system (endolymphatic hydrops) leading to degeneration of vestibular and cochlear hair cells. This may result from endolymphatic sac dysfunction secondary to infection, trauma, autoimmune disease, inflammatory causes, or tumor; an idiopathic etiology constitutes the largest category and is most accurately referred to as Ménière's disease. Endolymphatic sac tumors, often associated with von Hippel Lindau disease, may clinically mimic Ménière's disease. Although any pattern of hearing loss can be observed, typically, low-frequency, unilateral sensorineural hearing impairment is present. An abnormal VEMP test (see below) may be helpful in detecting Ménière's disease in a clinically unaffected contralateral ear. MRI should be obtained to exclude retrocochlear pathology such as a cerebellopontine angle tumor, endolymphatic sac tumor, or demyelinating disorder. Therapy is directed toward the control of vertigo. A 2-g/d low-salt diet is the mainstay of treatment for control of rotatory vertigo. Diuretics, a short course of oral glucocorticoids, intratympanic glucocorticoids, or intratympanic gentamicin may also be useful adjuncts in recalcitrant cases. Surgical therapy of vertigo is reserved for unresponsive cases and includes endolymphatic sac decompression, labyrinthectomy, and vestibular nerve section. Both labyrinthectomy and vestibular nerve section abolish rotatory vertigo in >90% of cases. Unfortunately, there is no effective therapy for hearing loss, tinnitus, or aural fullness from Ménière's disease.

Sensorineural hearing loss may also result from any neoplastic, vascular, demyelinating, infectious, degenerative disease, or trauma affecting the central auditory pathways. Characteristically, in hearing loss due to central nervous system pathology, a reduction in clarity of hearing and speech comprehension is much greater than the loss of the ability to hear pure tone. Auditory testing is consistent with an auditory neuropathy; normal otoacoustic emissions (OAEs) and an abnormal auditory brainstem response (ABR) are typical (see below). Hearing loss can accompany hereditary sensorimotor neuropathies and inherited disorders of myelin. Tumors of the cerebellopontine angle such as vestibular schwannoma and meningioma **(Chap. 90)** usually present with asymmetric sensorineural hearing loss with greater deterioration of speech understanding than pure tone hearing. Multiple sclerosis **(Chap. 444)** may present with acute unilateral or bilateral hearing loss; typically, pure tone testing remains relatively stable while speech understanding fluctuates. Isolated labyrinthine infarction can present with acute hearing loss and vertigo due to a cerebrovascular accident involving the posterior circulation, usually the anterior inferior cerebellar artery; it may also be the heralding sign of impending catastrophic basilar artery infarction **(Chap. 426)**. HIV **(Chap. 202)**, which can produce both peripheral and central auditory system pathology, is another consideration in the evaluation of sensorineural hearing impairment.

A finding of conductive and sensorineural hearing loss in combination is termed *mixed hearing loss*. Mixed hearing losses can result from pathology of both the middle and inner ear, as can occur in otosclerosis involving the ossicles and the cochlea, head trauma, chronic otitis media, cholesteatoma, middle-ear tumors, and some inner-ear malformations.

Trauma resulting in temporal bone fractures may be associated with conductive, sensorineural, or mixed hearing loss. If the fracture spares the inner ear, there may simply be conductive hearing loss due to rupture of the tympanic membrane or disruption of the ossicular chain. These abnormalities can be surgically corrected. Profound hearing loss and severe vertigo are associated with temporal bone fractures involving the inner ear. A perilymphatic fistula associated with leakage of inner-ear fluid into the middle ear can occur and may require surgical repair. An associated facial nerve injury is not uncommon. CT is best suited to assess fracture of the traumatized temporal bone, evaluate the ear canal, and determine the integrity of the ossicular chain and involvement of the inner ear. Cerebrospinal fluid leaks that accompany temporal bone fractures are usually self-limited; the value of prophylactic antibiotics is uncertain.

Tinnitus

Tinnitus is defined as the perception of a sound when there is no sound in the environment. It can have a buzzing, roaring, or ringing quality and may be pulsatile (synchronous with the heartbeat). Tinnitus is often associated with either a conductive or sensorineural hearing loss. The pathophysiology of tinnitus is not well understood. The cause of the tinnitus can usually be determined by finding the cause of the associated hearing loss. Tinnitus may be the first symptom of a serious condition such as a vestibular schwannoma. Pulsatile tinnitus requires evaluation of the vascular system of the head to exclude vascular tumors such as glomus jugulare tumors, aneurysms, dural arteriovenous fistulas, and stenotic arterial lesions; it may also occur with SOM, superior semicircular dehiscence, and inner-ear dehiscence. It is most commonly associated with some abnormality of the jugular bulb such as a large jugular bulb or jugular bulb diverticulum. In absence of demonstrated pathology on MRA/MRV or CT angiography, pulsatile tinnitus is usually attributed to turbulent venous blood flow through the transverse sinus, sigmoid sinus, and the jugular bulb.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

More than half of childhood hearing impairment is thought to be hereditary; hereditary hearing impairment (HHI) can also manifest later in life. HHI

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

as a large jugular bulb or jugular bulb diverticulum. In absence of demonstrated pathology on MRA/MRV or CT angiography, pulsatile tinnitus is usually attributed to turbulent venous blood flow through the transverse sinus, sigmoid sinus, and the jugular bulb.

GENETIC CAUSES OF HEARING LOSS



More than half of childhood hearing impairment is thought to be hereditary; hereditary hearing impairment (HHI) can also manifest later in life. HHI may be classified as either nonsyndromic, when hearing loss is the only clinical abnormality, or syndromic, when hearing loss is associated with anomalies in other organ systems. Nearly two-thirds of HHIs are nonsyndromic. Between 70% and 80% of nonsyndromic HHI is inherited in an autosomal recessive manner and designated DFNB; another 15–20% is autosomal dominant (DFNA). Less than 5% is X-linked (DFNX) or maternally inherited via the mitochondria.

More than 150 loci harboring genes for nonsyndromic HHI have been mapped, with recessive loci outnumbering dominant ones; numerous genes have now been identified **(Table 34-1)**. The hearing genes fall into the categories of structural proteins (*MYH9, MYO7A, MYO15, TECTA, DIAPH1*), transcription factors (*POU3F4, POU4F3*), ion channels (*KCNQ4, SLC26A4*), and gap junction proteins (*GJB2, GJB3, GJB6*). Several of these genes, including *GJB2, TECTA*, and *TMC1*, cause both autosomal dominant and recessive forms of nonsyndromic HHI. In general, the hearing loss associated with dominant genes has its onset in adolescence or adulthood, varies in severity, and progresses with age, whereas the hearing loss associated with recessive inheritance is congenital and profound. Connexin 26, a product of the *GJB2* gene, is particularly important because it is responsible for nearly 20% of all cases of childhood deafness; half of genetic deafness in children is *GJB2* related. Two frameshift mutations, 35delG and 167delT, account for >50% of the cases; however, screening for these two mutations alone is insufficient, and sequencing of the entire gene is required to fully capture *GJB2*-related recessive deafness. The 167delT mutation is highly prevalent in Ashkenazi Jews; ~1 in 1765 individuals in this population is homozygous and affected. *GJB2* hearing loss can also vary among the members of the same family, suggesting that other genes or factors influence the auditory phenotype. A single mutation in *GJB2* in combination with a single mutation in *GJB6* (connexin 30) can also lead to hearing loss and is an example of digenic inheritance of hearing loss.

TABLE 34-1

Hereditary Hearing Impairment Genes

DESIGNATION	GENE	FUNCTION
Autosomal Dominant		
DFNA1	*DIAPH1*	Cytoskeletal protein
DFNA2A	*KCNQ4*	Potassium channel
DFNA2B	*GJB3*	Gap junction
DFNA2C	*IFNLR1*	Class II cytokine receptor
DFNA3A	*GJB2*	Gap junction
DFNA3B	*GJB6*	Gap junction
DFNA4A	*MYH14*	Class II nonmuscle myosin
DFNA4B	*CEACAM16*	Cell adhesion molecule
DFNA5	*GSDME/DFNA5*	Executioner of pyroptosis
DFNA6/14/38	*WFS1*	Transmembrane protein
DFNA7	*LMX1A*	Transcription factor
DFNA9	*COCH*	Unknown

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

DFNA7	*LMX1A*	Transcription factor
DFNA8/12	*TECTA*	Tectorial membrane protein
DFNA9	*COCH*	Unknown
DFNA10	*EYA4*	Developmental gene
DFNA11	*MYO7A*	Cytoskeletal protein
DFNA13	*COL11A2*	Cytoskeletal protein
DFNA15	*POU4F3*	Transcription factor
DFNA17	*MYH9*	Cytoskeletal protein
DFNA20/26	*ACTG1*	Cytoskeletal protein
DFNA22	*MYO6*	Unconventional myosin
DFNA23	*SIX1*	Developmental gene
DFNA25	*SLC17A8*	Vesicular glutamate transporter
DFNA27	*REST*	Transcriptional repressor
DFNA28	*GRHL2*	Transcription factor
DFNA34	*NLRP3*	Pyrin-like protein involved in inflammation
DFNA36	*TMC1*	Transmembrane protein
DNA37	*COL11A1*	Cytoskeletal protein
DFNA40	*CRYM*	Thyroid hormone–binding protein
DFNA41	*P2RX2*	Purinergic receptor
DFNA44	*CCDC50*	Effector of epidermal growth factor–mediated signaling
DFNA50	*MIRN96*	MicroRNA
DFNA51	*TJP2*	Tight junction protein
DFNA56	*TNC*	Extracellular matrix protein
DFNA64	*SMAC/DIABLO*	Mitochondrial proapoptotic protein
DFNA65	*TBC1D24*	ARF6-interacting protein
DFNA66	*CD164*	Sialomucin
DFNA67	*OSBPL2*	Intracellular lipid receptor
DFNA69	*KIT LG*	Ligand for KIT receptor

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

DFNA67	*OSBPL2*	Intracellular lipid receptor
DFNA68	*HOMER2*	Stereociliary scaffolding protein
DFNA69	*KITLG*	Ligand for KIT receptor
DFNA70	*MCM2*	Initiation and elongation during DNA replication
DFNA73	*PTPRQ*	Member of type III receptor-like protein-tyrosine phosphatase (PTPase) family
	DMXL2	Regulator of Notch signaling
	MYO3A	Member of myosin superfamily
	PDE1C	Catalyze hydrolysis of cAMP and cGMP
	TRRAP	Transformation/transcription domain associated protein
	PLS1	Actin-bundling protein
	SCD5	Catalyzes formation of monounsaturated fatty acids from saturated fatty acids
	SLC12A2	Sodium-potassium-chloride transporter
	MAP1B	Microtubule binding protein
	RIPOR2/FAM65B	Membrane-associated protein in stereocilia
Autosomal Recessive		
DFNB1A	*GJB2*	Gap junction
DFNB1B	*GJB6*	Gap junction
DFNB2	*MYO7A*	Cytoskeletal protein
DFNB3	*MYO15A*	Cytoskeletal protein
DFNB4	*SLC26A4*	Chloride/iodide transporter
DFNB6	*TMIE*	Transmembrane protein
DFNB7/B11	*TMC1*	Transmembrane protein
DFNB8/10	*TMPRSS3*	Transmembrane serine protease
DFNB9	*OTOF*	Trafficking of membrane vesicles
DFNB12	*CDH23*	Intercellular adherence protein
DFNB15/72/95	*GIPC3*	PDZ domain–containing protein
DFNB16	*STRC*	Stereocilia protein
DFNB18B	*OTOG*	Tectorial membrane protein

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

DFNB16	*STRC*	Stereocilia protein
DFNB18	*USH1C*	Unknown
DFNB18B	*OTOG*	Tectorial membrane protein
DFNB21	*TECTA*	Tectorial membrane protein
DFNB22	*OTOA*	Gel attachment to nonsensory cell
DFNB23	*PCDH15*	Morphogenesis and cohesion
DFNB24	*RDX*	Cytoskeletal protein
DFNB25	*GRXCR1*	Reversible S-glutathionylation of proteins
DFNB26	*GAB1*	Member of insulin receptor substrate 1–like multisubstrate docking adapter protein family
DFNB28	*TRIOBP*	Cytoskeletal-organizing protein
DFNB29	*CLDN14*	Tight junctions
DFNB30	*MYO3A*	Hybrid motor-signaling myosin
DFNB31	*WHRN*	PDZ domain–containing protein
DFNB32/105	*CDC14A*	Protein phosphatase involved in hair cell ciliogenesis
DFNB35	*ESRRB*	Estrogen-related receptor beta protein
DFNB36	*ESPN*	Ca-insensitive actin-bundling protein
DFNB37	*MYO6*	Unconventional myosin
DFNB39	*HFG*	Hepatocyte growth factor
DFNB42	*ILDR1*	Ig-like domain–containing receptor
DFNB44	*ADCY1*	Adenylate cyclase
DFNB48	*CIB2*	Calcium and integrin binding protein
DFNB49	*BDP1*	Subunit of RNA polymerase
DFNB49	*MARVELD2*	Tight junction protein
DFNB53	*COL11A2*	Collagen protein
DFNB59	*PJVK*	Zn-binding protein
DFNB60	*SLC22A4*	Prestin, motor protein of cochlear outer hair cell
DFNB61	*SLC26A5*	Motor protein
DFNB66	*DCDC2*	Ciliary protein

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

DFNB61	*SLC26A5*	Motor protein
DFNB63	*LRTOMT/COMT2*	Putative methyltransferase
DFNB66	*DCDC2*	Ciliary protein
DFNB66/67	*LHFPL5*	Tetraspan protein
DFNB68	*S1PR2*	Tetraspan membrane protein of hair cell stereocilia
DFNB70	*PNPT1*	Mitochondrial–RNA–import protein
DFNB73	*BSND*	Beta subunit of chloride channel
DFNB74	*MSRB3*	Methionine sulfoxide reductase
DFNB76	*SYNE4*	Part of *LINC* tethering complex
DFNB77	*LOXHD1*	Stereociliary protein
DFNB79	*TPRN*	Unknown
DFNB82	*GPSM2*	G protein signaling modulator
DFNB84	*PTPRQ*	Type III receptor-like protein-tyrosine phosphatase family
DFNB84	*OTOGL*	Otogelin-like protein
DFNB86	*TBC1D24*	GTPase-activating protein
DFNB88	*ELMOD3*	GTPase-activating protein
DFNB89	*KARS*	Lysyl-tRNA synthetase
DFNB91	*SERPINB6*	Protease inhibitor
DFNB93	*CABP2*	Calcium-binding protein
DFN94	*NARS2*	Mitochondrial asparaginyl-tRNA synthetase
DFNA97	*MET*	Oncogene/hepatocyte growth factor receptor
DFNB98	*TSPEAR*	Epilepsy-associated repeats containing protein
DFNB99	*TMEM132E*	Transmembrane protein
DFNB100	*PPIP5K2*	Diphosphoinositol-pentakisphosphate kinase
DFNB101	*GRXCR2*	Maintaining stereocilia bundles
DFNB102	*EPS8*	Epidermal growth factor receptor
DFNB103	*CLIC5*	Chloride ion transport
DFNB106	*EPS8L2*	Actin remodeling in response to EGF stimulation

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

DFNB103	*CLIC5*	Chloride ion transport
DFNB104	*FAM65B/RIPOR2*	Membrane-associated protein in stereocilia
DFNB106	*EPS8L2*	Actin remodeling in response to EGF stimulation
DFNB108	*ROR1*	Receptor tyrosine kinase-like orphan receptor
	WBP2	Transcriptional coactivator for estrogen receptor-alpha and progesterone receptor
	ESRP1	Modulates activation of G proteins
	MPZL2	Mediates epithelial cell-cell interactions in developing tissues
	CEACAM16	Cell adhesion molecule
	GRAP	Cytoplasmic signaling protein
	SPNS2	Sphingosine-1-phosphate (S1P) transporter
	CLDN9	Tight junctions
	CLRN2	Maintenance of transducing stereocilia in auditory hair cells
X-linked		
DFNX1	PRPS1	Catalyzes phosphoribosylation of ribose 5-phosphate to 5-phosphoribosyl-1-pyrophosphate
DFNX2	POU3F4	Transcription factor
DFNX4	SMPX	Small muscle protein
DFNX5	AIFM1	Mitochondrial flavin adenine dinucleotide (FAD)-dependent oxidoreductase
DFNX6	COL4A6	Collagen protein

In addition to *GJB2*, several other nonsyndromic genes are associated with hearing loss that progresses with age. The contribution of genetics to presbycusis is also becoming better understood and likely reflects a combination of genetic susceptibility impacted by environmental exposure to sound. Sensitivity to aminoglycoside ototoxicity can be maternally transmitted through a mitochondrial mutation. Susceptibility to noise-induced hearing loss may also be genetically determined.

There are >400 syndromic forms of hearing loss. These include Usher's syndrome (retinitis pigmentosa and hearing loss), Waardenburg's syndrome (pigmentary abnormality and hearing loss), Pendred's syndrome (thyroid organification defect and hearing loss), Alport's syndrome (renal disease and hearing loss), Jervell and Lange-Nielsen syndrome (prolonged QT interval and hearing loss), neurofibromatosis type 2 (bilateral acoustic schwannoma), and mitochondrial disorders (mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes [MELAS]; myoclonic epilepsy and ragged red fibers [MERRF]; and progressive external ophthalmoplegia [PEO]) **(Table 34-2)**.

TABLE 34-2

Syndromic Hereditary Hearing Impairment Genes

SYNDROME	GENE	FUNCTION
BOR syndrome	*EYA1*	Developmental gene

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

SYNDROME	GENE	FUNCTION
Alport's syndrome	*COL4A3-5*	Cytoskeletal protein
BOR syndrome	*EYA1*	Developmental gene
	SIX5	Developmental gene
	SIX1	Developmental gene
Jervell and Lange-Nielsen syndrome	*KCNQ1* *KCNE1*	Delayed rectifier K^+ channel Delayed rectifier K^+ channel
Norrie's disease	*NDP*	Cell–cell interactions
Pendred's syndrome	*SLC26A4*	Chloride/iodide transporter
	FOXI1 *KCNJ10*	Transcriptional activator of *SLC26A4* Inwardly rectifying K^+ channel
Treacher Collins syndrome	*TCOF1* *POLR1D* *POLR1C*	Nucleolar-cytoplasmic transport Subunit of RNA polymerases I and III Subunit of RNA polymerases I and III
Usher's syndrome	*MYO7A*	Cytoskeletal protein
	USH1C	Unknown
	CDH23	Intercellular adherence protein
	PCDH15	Cell adhesion molecule
	SANS *CIB2*	Harmonin-associated protein Calcium- and integrin-binding protein
	USH2A	Cell adhesion molecule
	VLGR1	G protein–coupled receptor
	WHRN	PDZ domain–containing protein
	CLRN1 *HARS* *PDZD7*	Cellular synapse protein Histidyl-tRNA synthetase PDZ domain–containing protein
WS type I, III	*PAX3*	Transcription factor
WS type II	*MITF*	Transcription factor
	SNAI2	Transcription factor
WS type IV	*EDNRB*	Endothelin B receptor

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

	EDN3	Endothelin B receptor ligand
	SOX10	Transcription factor

Abbreviations: BOR, branchio-oto-renal syndrome; WS, Waardenburg's syndrome.

APPROACH TO THE PATIENT

Disorders of the Sense of Hearing

The goal in the evaluation of a patient with auditory complaints is to determine (1) the nature of the hearing impairment (conductive vs sensorineural vs mixed), (2) the severity of the impairment (mild, moderate, severe, or profound), (3) the anatomy of the impairment (external ear, middle ear, inner ear, or central auditory pathway), and (4) the etiology. The presence of signs and symptoms associated with hearing loss should be ascertained **(Table 34-3)**. The history should elicit characteristics of the hearing loss, including the duration of deafness, unilateral versus bilateral involvement, nature of onset (sudden vs insidious), and rate of progression (rapid vs slow). Symptoms of tinnitus, vertigo, imbalance, aural fullness, otorrhea, headache, facial nerve dysfunction, and head and neck paresthesias should be noted. Information regarding head trauma, exposure to ototoxins, occupational or recreational noise exposure, and family history of hearing impairment may also be important. A sudden onset of unilateral hearing loss, with or without tinnitus, may represent a viral infection of the inner ear, vestibular schwannoma, or a stroke. Patients with unilateral hearing loss (sensory or conductive) usually complain of reduced hearing, poor sound localization, and difficulty hearing clearly in the presence of background noise. Gradual progression of a hearing deficit is common with otosclerosis, noise-induced hearing loss, vestibular schwannoma, or Ménière's disease. Small vestibular schwannomas typically present with asymmetric hearing impairment, tinnitus, and imbalance (rarely vertigo); cranial neuropathy, in particular of the trigeminal or facial nerve, may accompany larger tumors. In addition to hearing loss, Ménière's disease may be associated with episodic vertigo, tinnitus, and aural fullness. Sound-induced vertigo, autophony, and being able to hear one's own neck or eye movement are highly suggestive of superior semicircular canal dehiscence. Hearing loss with otorrhea is most likely due to chronic otitis media or cholesteatoma.

Examination should include the auricle, external ear canal, and tympanic membrane. In the elderly, the external ear canal is often dry and fragile; it is preferable to clean cerumen with wall-mounted suction or cerumen loops and to avoid irrigation. Irrigation should also be avoided when a tympanic membrane perforation is present or the integrity of the eardrum cannot be established. In examining the eardrum, the topography of the tympanic membrane is more important than the presence or absence of the light reflex. In addition to the pars tensa (the lower two-thirds of the tympanic membrane), the pars flaccida (upper one-third of the tympanic membrane) above the short process of the malleus should also be examined for retraction pockets that may be evidence of chronic Eustachian tube dysfunction or cholesteatoma. Insufflation of the ear canal is necessary to assess tympanic membrane mobility and compliance. Careful inspection of the nose, nasopharynx, and upper respiratory tract is important. Unilateral serous effusion or unexplained otalgia should prompt a fiberoptic examination of the nasopharynx and larynx to exclude neoplasms. Cranial nerves should be evaluated with special attention to facial and trigeminal nerves, which are commonly affected with tumors involving the cerebellopontine angle.

The Rinne and Weber tuning fork tests, with a 512-Hz tuning fork, are used to screen for hearing loss, differentiate conductive from sensorineural hearing losses, and confirm the findings of audiologic evaluation. The Rinne test compares the ability to hear by air conduction with the ability to hear by bone conduction. The tines of a vibrating tuning fork are held near the opening of the external auditory canal, and then the stem is placed on the mastoid process; for direct contact, it may be placed on teeth or dentures. The patient is asked to indicate whether the tone is louder by air conduction or bone conduction. Normally, and in the presence of sensorineural hearing loss, a tone is heard louder by air conduction than by bone conduction; however, with conductive hearing loss of ≥30 dB (see "Audiologic Assessment," below), the bone-conduction stimulus is perceived as louder than the air-conduction stimulus. For the Weber test, the stem of a vibrating tuning fork is placed on the head in the midline and the patient is asked whether the tone is heard in both ears or better in one ear than in the other. With a unilateral conductive hearing loss, the tone is perceived in the affected ear. With a unilateral sensorineural hearing loss, the tone is perceived in the unaffected ear. A 5-dB difference in hearing between the two ears is required for lateralization.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Saying "huh" a great deal

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 34-3

Signs and Symptoms Suggestive of Hearing Loss

Saying “huh” a great deal
Reduced clarity of hearing
Difficulty understanding conversations in background noise
Family complaining of hearing loss
Tinnitus
Turning the volume up on radio or television
Sensitivity to noises
Fullness in the ear
Avoiding social settings

LABORATORY ASSESSMENT OF HEARING

Audiologic Assessment

The minimum audiologic assessment for hearing loss should include the measurement of pure tone air-conduction and bone-conduction thresholds, speech reception threshold, word recognition score, tympanometry, acoustic reflexes, and acoustic-reflex decay. This test battery provides a screening evaluation of the entire auditory system and allows one to determine whether further differentiation of a sensory (cochlear) from a neural (retrocochlear) hearing loss is indicated.

Pure tone audiometry assesses hearing acuity for pure tones. The test is administered by an audiologist and is performed in a sound-attenuated chamber. The pure tone stimulus is delivered with an audiometer, an electronic device that allows the presentation of specific frequencies (generally between 250–8000 Hz) at specific intensities. Air- and bone-conduction thresholds are established for each ear. Air-conduction thresholds are determined by presenting the stimulus in air with the use of headphones. Bone-conduction thresholds are determined by placing the stem of a vibrating tuning fork or an oscillator of an audiometer in contact with the head. In the presence of a hearing loss, broad-spectrum noise is presented to the nontest ear for masking purposes so that responses are based on perception from the ear under test.

The responses are measured in decibels (dBs). An *audiogram* is a plot of intensity in dBs of hearing threshold versus frequency. A dB is equal to 20 times the logarithm of the ratio of the sound pressure required to achieve threshold in the patient to the sound pressure required to achieve threshold in a normal-hearing person. Therefore, a change of 6 dB represents doubling of sound pressure, and a change of 20 dB represents a tenfold change in sound pressure. Loudness, which depends on the frequency, intensity, and duration of a sound, doubles with approximately each 10-dB increase in sound pressure level. Pitch, on the other hand, does not directly correlate with frequency. The perception of pitch changes slowly in the low and high frequencies. In the middle tones, which are important for human speech, pitch varies more rapidly with changes in frequency.

Pure tone audiometry establishes the presence and severity of hearing impairment, unilateral versus bilateral involvement, and the type of hearing loss. Conductive hearing losses with a large mass component, as is often seen in middle-ear effusions, produce elevation of thresholds that predominate in the higher frequencies. Conductive hearing losses with a large stiffness component, as in fixation of the footplate of the stapes in early otosclerosis, produce threshold elevations in the lower frequencies. Often, the conductive hearing loss involves all frequencies, suggesting involvement of both stiffness and mass. In general, sensorineural hearing losses such as presbycusis affect higher frequencies more than lower frequencies (Fig. 34-3). An exception is Ménière’s disease, which is characteristically associated with low-frequency sensorineural hearing loss (though

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Speech recognition requires greater synchronous neural firing than is necessary for appreciation of pure tones. *Speech audiometry* tests the clarity with which one hears. The *speech reception threshold* (*SRT*) is defined as the intensity at which speech is recognized as a meaningful symbol and is

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

involvement of both stiffness and mass. In general, sensorineural hearing losses such as presbycusis affect higher frequencies more than lower frequencies (Fig. 34-3). An exception is Ménière's disease, which is characteristically associated with low-frequency sensorineural hearing loss (though any frequency can be affected). Noise-induced hearing loss has an unusual pattern of hearing impairment in which the loss at 3000–4000 Hz is greater than at higher frequencies. Vestibular schwannomas characteristically affect the higher frequencies, but any pattern of hearing loss can be observed.

Speech recognition requires greater synchronous neural firing than is necessary for appreciation of pure tones. *Speech audiometry* tests the clarity with which one hears. The *speech reception threshold* (*SRT*) is defined as the intensity at which speech is recognized as a meaningful symbol and is obtained by presenting two-syllable words with an equal accent on each syllable. The intensity at which the patient can repeat 50% of the words correctly is the SRT. Once the SRT is determined, discrimination or word recognition ability is tested by presenting one-syllable words at 25–40 dB above the SRT. The words are phonetically balanced in that the phonemes (speech sounds) occur in the list of words at the same frequency that they occur in ordinary conversational English. An individual with normal hearing or conductive hearing loss can repeat 88–100% of the phonetically balanced words correctly. Patients with a sensorineural hearing loss have variable loss of discrimination. As a general rule, neural lesions produce greater deficits in discrimination than do cochlear lesions. For example, in a patient with mild asymmetric sensorineural hearing loss, a clue to the diagnosis of vestibular schwannoma is the presence of greater than expected deterioration in discrimination ability. Deterioration in discrimination ability at higher intensities above the SRT also suggests a lesion in the eighth nerve or central auditory pathways.

Tympanometry measures the impedance of the middle ear to sound and is useful in diagnosis of middle-ear effusions. A *tympanogram* is the graphic representation of change in impedance or compliance as the pressure in the ear canal is changed. Normally, the middle ear is most compliant at atmospheric pressure, and the compliance decreases as the pressure is increased or decreased (type A); this pattern is seen with normal hearing or in the presence of sensorineural hearing loss. Compliance that does not change with change in pressure suggests middle-ear effusion (type B). With a negative pressure in the middle ear, as with Eustachian tube obstruction, the point of maximal compliance occurs with negative pressure in the ear canal (type C). A tympanogram in which no point of maximal compliance can be obtained is most commonly seen with discontinuity of the ossicular chain (type A_d). A reduction in the maximal compliance peak can be seen in otosclerosis (type A_s).

During tympanometry, an intense tone elicits contraction of the stapedius muscle. The change in compliance of the middle ear with contraction of the stapedius muscle can be detected. The presence or absence of this *acoustic reflex* is important in determining the etiology of hearing loss as well as in the anatomic localization of facial nerve paralysis. The acoustic reflex can help differentiate between conductive hearing loss due to otosclerosis and that caused by an inner-ear "third window": it is absent in otosclerosis and present in inner-ear conductive hearing loss. Normal or elevated acoustic reflex thresholds in an individual with sensorineural hearing impairment suggest a cochlear hearing loss. An absent acoustic reflex in the setting of sensorineural hearing loss is not helpful in localizing the site of lesion. Assessment of *acoustic reflex decay* helps differentiate sensory from neural hearing losses. In neural hearing loss, such as with vestibular schwannoma, the reflex adapts or decays with time.

OAEs generated by outer hair cells only can be measured with microphones inserted into the external auditory canal. The emissions may be spontaneous or evoked with sound stimulation. The presence of OAEs indicates that the outer hair cells of the organ of Corti are intact and can be used to assess auditory thresholds and to distinguish sensory from neural hearing losses.

Evoked Responses

Electrocochleography measures the earliest evoked potentials generated in the cochlea and the auditory nerve. Receptor potentials recorded include the cochlear microphonic, generated by the outer hair cells of the organ of Corti, and the summating potential, generated by the inner hair cells in response to sound. The whole nerve action potential representing the composite firing of the first-order neurons can also be recorded during electrocochleography. Clinically, the test is useful in the diagnosis of Ménière's disease, in which an elevation of the ratio of summating potential to action potential is seen.

Brainstem auditory-evoked responses (*BAERs*), also known as ABRs, are useful in differentiating the site of sensorineural hearing loss. In response to sound, five distinct electrical potentials arising from different stations along the peripheral and central auditory pathway (eighth nerve, cochlear nucleus, superior olivary complex, lateral lemniscus, and inferior colliculus) can be identified using computer averaging from scalp surface electrodes. BAERs are valuable in situations in which patients cannot or will not give reliable voluntary thresholds. They are also used to assess the integrity of the auditory nerve and brainstem in various clinical situations, including intraoperative monitoring, and in determination of brain death.

The *VEMP test* investigates otolith and vestibular nerve function by presenting a high-level acoustic stimulus and evoking a short-latency electromyographic potential; cVEMP (or cervical VEMP) and oVEMP (or ocular VEMP) have been described. The cVEMP elicits a vestibulocollic reflex whose afferent limb arises from acoustically sensitive cells in the saccule, with signals conducted via the inferior vestibular nerve. cVEMP is a biphasic,

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

window"), and perilymphatic fistula. The oVEMP, in contrast, is a response involving the utricle primarily and superior vestibular nerve. The oVEMP excitatory response is recorded from the extraocular muscle. The oVEMP is abnormal in superior vestibular neuritis

electromyographic potential; cVEMP (or cervical VEMP) and oVEMP (or ocular VEMP) have been described. The cVEMP elicits a vestibulocollic reflex whose afferent limb arises from acoustically sensitive cells in the saccule, with signals conducted via the inferior vestibular nerve. cVEMP is a biphasic, short-latency response recorded from the tonically contracted sternocleidomastoid muscle in response to loud auditory clicks or tones. cVEMPs may be diminished or absent in patients with early and late Ménière's disease, vestibular neuritis, benign paroxysmal positional vertigo, and vestibular schwannoma. On the other hand, the threshold for VEMPs may be lower in cases of superior canal dehiscence, other inner-ear dehiscence ("third window"), and perilymphatic fistula. The oVEMP, in contrast, is a response involving the utricle primarily and superior vestibular nerve. The oVEMP excitatory response is recorded from the extraocular muscle. The oVEMP is abnormal in superior vestibular neuritis.

Imaging Studies

The choice of radiologic tests is largely determined by whether the goal is to evaluate the bony anatomy of the external, middle, and inner ear or to image the auditory nerve and brain. Axial and coronal CT of the temporal bone with fine 0.3-mm cuts is ideal for determining the caliber of the external auditory canal, integrity of the ossicular chain, and presence of middle-ear or mastoid disease; it can also detect inner-ear malformations. CT is also ideal for the detection of bone erosion with chronic otitis media and cholesteatoma. Pöschl reformatting in the plane of the superior semicircular canal is required for the identification of dehiscence or absence of bone over the superior semicircular canal. MRI is superior to CT for imaging of retrocochlear pathology such as vestibular schwannoma, meningioma, other lesions of the cerebellopontine angle, demyelinating lesions of the brainstem, and brain tumors. Both CT and MRI are equally capable of identifying inner-ear malformations and assessing cochlear patency for preoperative evaluation of patients for cochlear implantation.

TREATMENT

Disorders of the Sense of Hearing

In general, conductive hearing losses are amenable to surgical correction, whereas sensorineural hearing losses are usually managed medically. Atresia of the ear canal can be surgically repaired, often with significant improvement in hearing. Alternatively, the conductive hearing loss associated with atresia can be addressed with a bone-anchored hearing aid (BAHA). Tympanic membrane perforations due to chronic otitis media or trauma can be repaired with an outpatient tympanoplasty. Likewise, conductive hearing loss associated with otosclerosis can be treated by stapedectomy, which is successful in >95% of cases. Tympanostomy tubes allow the prompt return of normal hearing in individuals with middle-ear effusions. Hearing aids are effective and well tolerated in patients with conductive hearing losses.

Patients with mild, moderate, and severe sensorineural hearing losses are regularly rehabilitated with hearing aids of varying configuration and strength. Hearing aids have been improved to provide greater fidelity and have been miniaturized. The current generation of hearing aids is nearly invisible, thus reducing stigma associated with their use. In general, the more severe the hearing impairment, the larger the hearing aid required for auditory rehabilitation. Digital hearing aids lend themselves to individual programming, and multiple and directional microphones at the ear level may be helpful in noisy surroundings. Because all hearing aids amplify noise as well as speech, the only absolute solution to the problem of noise is to place the microphone closer to the speaker than the noise source. This arrangement is not possible with a self-contained, cosmetically acceptable device. A significant limitation of rehabilitation with a hearing aid is that although it is able to enhance detection of sound with amplification, it cannot restore clarity of hearing that is lost with presbycusis.

The cost of a single hearing aid (~$2300 US) is a significant obstacle for many hearing-impaired individuals and usually bilateral amplification is recommended. To reduce cost and spur innovation, a new category of over-the-counter amplification devices that can be purchased similar to reading eyeglasses by simply walking into a store has recently been approved by the US Food and Drug Administration. By reducing the cost of amplification devices to consumers, promoting innovation, and increasing competition, this new class of devices could fundamentally change the way hearing rehabilitation is delivered.

Patients with unilateral deafness have difficulty with sound localization and reduced clarity of hearing in background noise. They may benefit from a contralateral routing of signal (CROS) hearing aid in which a microphone is placed on the hearing-impaired side, and the sound is transmitted to the receiver placed on the contralateral ear. The same result may be obtained with a BAHA, in which a hearing aid clamps to a screw integrated into the skull on the hearing-impaired side. Like the CROS hearing aid, the BAHA transfers the acoustic signal to the contralateral hearing ear, but it does so by vibrating the skull. Patients with profound deafness on one side and some hearing loss in the better ear are candidates for a BICROS hearing aid; it differs from the CROS hearing aid in that the patient wears a hearing aid, and not simply a receiver, in the better ear. Unfortunately, while CROS and BAHA devices provide benefit, they do not restore hearing in the deaf ear. Only cochlear implants can restore hearing (see below). Increasingly, cochlear implants are being used for the treatment of patients with single-sided deafness; they show great promise in not only restoring hearing

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

having the speaker closer to the microphone than any source of noise. Assistive devices include infrared and frequency-modulated (FM) transmission as well as an electromagnetic loop around the room for transmission to the individual's hearing aid. Hearing aids with telecoils can

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

and BAHA devices provide benefit, they do not restore hearing in the deaf ear. Only cochlear implants can restore hearing (see below). Increasingly, cochlear implants are being used for the treatment of patients with single-sided deafness; they show great promise in not only restoring hearing and reducing tinnitus, but also improving sound localization and performance in background noise.

In many situations, including lectures and the theater, hearing-impaired persons benefit from assistive devices that are based on the principle of having the speaker closer to the microphone than any source of noise. Assistive devices include infrared and frequency-modulated (FM) transmission as well as an electromagnetic loop around the room for transmission to the individual's hearing aid. Hearing aids with telecoils can also be used with properly equipped telephones in the same way. Bluetooth technology has revolutionized connectivity between hearing aids and other devices such as smart phones.

In the event that the hearing aid provides inadequate rehabilitation, cochlear implants may be appropriate **(Fig. 34-4)**. Criteria for implantation include severe to profound hearing loss with open-set sentence cognition of ≤40% under best-aided conditions. Worldwide, >600,000 hearing-impaired individuals have received cochlear implants. Cochlear implants are neural prostheses that convert sound energy to electrical energy and can be used to stimulate the auditory division of the eighth nerve directly. In most cases of profound hearing impairment, the auditory hair cells are lost but the ganglionic cells of the auditory division of the eighth nerve are preserved. Cochlear implants consist of electrodes that are inserted into the cochlea through the round window, speech processors that extract acoustic elements of speech for conversion to electrical currents, and a means of transmitting the electrical energy through the skin. Patients with implants experience sound that helps with speech reading, allows open-set word recognition, and helps in modulating the person's own voice. Usually, within the first 3–6 months after implantation, adult patients can understand speech without visual cues. With the current generation of multichannel cochlear implants, nearly 75% of patients are able to converse on the telephone. Bilateral cochlear implantations are commonly performed, especially in children; these patients perform better in background noise, have better sound localization, and are less fatigued by the "work" compared to monaural hearing.

Hybrid cochlear implants are indicated for the treatment of high-frequency hearing loss in patients who do not have profound hearing loss and yet do not benefit from hearing aids. Patients with presbycusis typically have normal low-frequency hearing while suffering from high-frequency hearing loss associated with loss of clarity that cannot always be adequately rehabilitated with a hearing aid. However, these patients are not candidates for conventional cochlear implants because they have too much residual hearing. The hybrid implant has been specifically designed for this patient population; it has a shorter electrode than a conventional cochlear implant and can be introduced into the cochlea atraumatically, thus preserving low-frequency hearing. Individuals with a hybrid implant use their own natural low-frequency "acoustic" hearing and rely on the implant for providing "electrical" high-frequency hearing. Patients who have received the hybrid implant perform better on speech discrimination tests in both quiet and noisy backgrounds.

For individuals who were born without cochlea or have had both eighth nerves destroyed by trauma or bilateral vestibular schwannomas (e.g., neurofibromatosis type 2), brainstem auditory implants placed near the cochlear nucleus may provide auditory rehabilitation. Currently, brainstem implants provide sound awareness but unfortunately speech understanding remains elusive.

Tinnitus often accompanies hearing loss. Similar to background noise, tinnitus can degrade speech comprehension in individuals with hearing impairment. Patients with tinnitus should be advised to minimize caffeine ingestion, avoid high dosage of nonsteroidal anti-inflammatory drugs (NSAIDs), and reduce stress. Therapy for tinnitus is usually directed toward minimizing the appreciation of tinnitus. Relief of the tinnitus may be obtained by masking it with background music or white noise. Hearing aids are also helpful in tinnitus suppression, as are tinnitus maskers, devices that present a sound to the affected ear that is more pleasant to listen to than the tinnitus. The use of a tinnitus masker is often followed by several hours of inhibition of the tinnitus. Antidepressants have also been shown to be beneficial in helping patients cope with tinnitus.

Hard-of-hearing individuals often benefit from a reduction in unnecessary noise in the environment (e.g., radio or television) to enhance the signal-to-noise ratio. Speech comprehension is aided by lip reading; therefore, the impaired listener should be seated so that the face of the speaker is well illuminated and easily seen. Although speech should be in a loud, clear voice, one should be aware that in sensorineural hearing losses in general and in hard-of-hearing elderly in particular, recruitment (abnormal perception of loud sounds) may be troublesome. Above all, optimal communication cannot take place without both parties giving it their full and undivided attention.

FIGURE 34-4

A cochlear implant is composed of an external microphone and speech processor worn on the ear and a receiver implanted underneath the temporalis muscle. The internal receiver is attached to an electrode that is placed surgically in the cochlea.

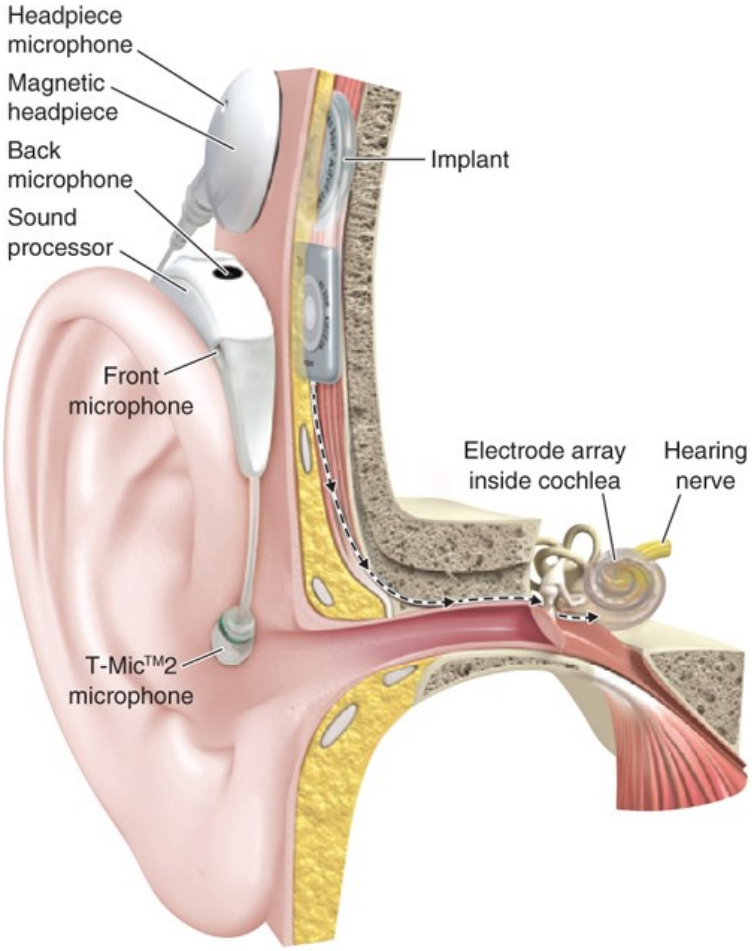
Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

A cochlear implant is composed of an external microphone and speech processor worn on the ear and a receiver implanted underneath the temporalis muscle. The internal receiver is attached to an electrode that is placed surgically in the cochlea.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

PREVENTION

Conductive hearing losses may be prevented by prompt antibiotic therapy of adequate duration for AOM and by ventilation of the middle ear with tympanostomy tubes in middle-ear effusions lasting ≥12 weeks. Loss of vestibular function and deafness due to aminoglycoside antibiotics can largely be prevented by careful monitoring of serum peak and trough levels.

Some 10 million Americans have noise-induced hearing loss, and 20 million are exposed to hazardous noise in their employment. Noise-induced hearing loss can be prevented by avoidance of exposure to loud noise or by regular use of earplugs or fluid-filled ear muffs to attenuate intense sound. Table 34-4 lists loudness levels for a variety of environmental sounds. High-risk activities for noise-induced hearing loss include use of electrical equipment for wood- and metalworking, and target practice or hunting with small firearms. All internal-combustion and electric engines, including snow and leaf blowers, snowmobiles, outboard motors, and chainsaws, require protection of the user with hearing protectors. Virtually all noise-induced hearing loss is preventable through education, which should begin before the teenage years. Programs for conservation of hearing in the workplace are required by the Occupational Safety and Health Administration (OSHA) whenever the exposure over an 8-h period averages 85 dB. OSHA mandates that workers in such noisy environments have hearing monitoring and protection programs that include a preemployment screen, an annual audiologic assessment, and the mandatory use of hearing protectors. Exposure to loud sounds above 85 dB in the work environment is restricted by OSHA, with halving of allowed exposure time for each increment of 5 dB above this threshold; for example, exposure to 90 dB is permitted for 8 h; 95 dB for 4 h, and 100 dB for 2 h (Table 34-5).

TABLE 34-4

Decibel (Loudness) Level of Common Environmental Noise

Weakest sound heard	0

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

for 8 h, 95 dB for 4 h, and 100 dB for 2 h (Table 34-5).

TABLE 34-4

Decibel (Loudness) Level of Common Environmental Noise

SOURCE	DECIBEL (dB)
Weakest sound heard	0
Whisper	30
Normal conversation	55–65
City traffic inside car	85
OSHA Monitoring Requirement Begins	**90**
Jackhammer	95
Subway train at 200 ft	95
Power mower	107
Power saw	110
Painful Sound	**125**
Jet engine at 100 ft	140
12-gauge shotgun blast	165
Loudest sound that can occur	194

Abbreviation: OSHA, Occupational Safety and Health Administration.

TABLE 34-5

OSHA Daily Permissible Noise Level Exposure

SOUND LEVEL (dB)	DURATION PER DAY (h)
90	8
92	6
95	4
97	3
100	2
102	1.5
115	≤0.25

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Abbreviation: OSHA, Occupational Safety and Health Administration.

TABLE 34-5

OSHA Daily Permissible Noise Level Exposure

SOUND LEVEL (dB)	DURATION PER DAY (h)
90	8
92	6
95	4
97	3
100	2
102	1.5
105	1
110	0.5
115	≤0.25

Note: Exposure to impulsive or impact noise should not exceed 140-dB peak sound pressure level.

Source: From *https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=standards&p_id=9735*.

FURTHER READING

Carlson ML: Cochlear implantation in adults. N Engl J Med 382:1531, 2020. [PubMed: 32294347]

Espinosa-Sanchez JM, Lopez-Escamez JA: Menière's disease. Handb Clin Neurol 137:257, 2016. [PubMed: 27638077]

Moser T, Starr A: Auditory neuropathy—neural and synaptic mechanisms. Nat Rev Neurol 12:135, 2016. [PubMed: 26891769]

Patel M et al: Intratympanic methylprednisolone versus gentamicin in patients with unilateral Ménière's disease: a randomised, double-blind, comparative effectiveness trial. Lancet 388:2753, 2016. [PubMed: 27865535]

Tikka C et al: Interventions to prevent occupational noise-induced hearing loss. Cochrane Database Syst Rev 7:CD006396, 2017. [PubMed: 28685503]

Wilson BS et al: Global hearing health care: new findings and perspectives. Lancet 390:2503, 2017. [PubMed: 28705460]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 34: Disorders of Hearing, Anil K. Lalwani
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 35: Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat

Rachel L. Amdur; Jeffrey A. Linder

INTRODUCTION

Upper respiratory symptoms are most commonly caused by viral infection but also can be caused by other infectious, inflammatory, allergic, autoimmune, and neoplastic conditions. This chapter will discuss ambulatory antibiotic prescribing and review the most common causes of upper respiratory symptoms, including nonspecific upper respiratory infections.

Ear pain is most commonly caused by otitis externa, acute otitis media (AOM), otitis media with effusion (OME), and acute mastoiditis. Sinus symptoms can be caused by acute sinusitis, invasive fungal sinusitis, nosocomial sinusitis, and chronic sinusitis. Sore throat and neck pain can be caused by streptococcal pharyngitis, nonstreptococcal pharyngitis, acute infectious mononucleosis, other types of bacterial pharyngitis, Lemierre's syndrome, gonococcal pharyngitis, diphtheria, acute HIV infection, head and neck abscesses, epiglottitis, and laryngitis. At the time of presentation, upper respiratory symptoms of most common viral and bacterial etiologies have generally lasted from hours up to a few days.

UPPER RESPIRATORY INFECTIONS

Upper respiratory infections (URIs) are acute respiratory infections that occur above the vocal cords. URIs, including nonspecific upper respiratory tract infection, otitis media, sinusitis, and pharyngitis, are collectively the most common symptomatic reason for seeking care in the United States. In terms of etiology, symptoms, and signs, URIs overlap with lower acute respiratory infections that occur below the vocal cords, such as influenza **(Chap. 200)**, acute bronchitis, and pneumonia **(Chap. 126)**, as well as with noninfectious cough **(Chap. 38)**. The average adult has 2–4 URIs per year; children can have 6–10 URIs annually. URIs can be prevented by hand washing or sanitization, physical distancing, use of facial masks, isolation of persons who are ill, and environmental cleaning **(Chap. 199)**.

SARS-CoV-2, the pathogen that causes COVID-19, can cause virtually any upper respiratory symptom **(Chap. 199)**. COVID-19 symptoms appear 2–14 days after exposure and may include fever, chills, cough, shortness of breath, fatigue, myalgias, headaches, rhinorrhea, sore throat, nausea, vomiting, or diarrhea. New loss of taste or smell appears to be specific for COVID-19. Until there is widespread natural or vaccine-induced immunity, any respiratory symptom occurring in areas where SARS-CoV-2 is circulating should be considered a potential manifestation of COVID-19.

IMPROVING AMBULATORY ANTIBIOTIC PRESCRIBING

The only common acute respiratory infections that should be treated with antibiotics are AOM, sinusitis, streptococcal pharyngitis, and pneumonia. Even for AOM, sinusitis, and pharyngitis, only a minority of cases meet the criteria for antibiotic prescribing. Common respiratory viruses **(Chap. 199)** cause the overwhelming majority of acute respiratory infections, and these infections are generally self-limited; antibiotics neither speed resolution nor prevent complications for the majority of acute respiratory infections. Unfortunately, for this reason, at least half of ambulatory antibiotic prescriptions for acute respiratory infections in the United States are inappropriate. Internationally, population rates of antibiotic prescribing vary nearly threefold, with no differences in infectious complications. Antibiotics cause adverse drug effects, alter the microbiome, cause *Clostridioides difficile* infection **(Chap. 134)**, increase health care costs, and increase the prevalence of antibiotic-resistant bacteria **(Chap. 145)**.

Clinicians prescribe inappropriate antibiotics because of time pressure; fear of missing a rare bacterial diagnosis; concern about preventing a rare bacterial complication; a lack of salience of adverse antibiotic effects; or a mistaken belief that most patients expect, demand, or will not be satisfied without an antibiotic prescription.

AMBULATORY ANTIBIOTIC STEWARDSHIP

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

improve antibiotic prescribing and assessing its effectiveness; (3) monitoring antibiotic prescribing and providing feedback; and (4) providing

Clinicians prescribe inappropriate antibiotics because of time pressure, fear of missing a rare bacterial diagnosis, concern about preventing a rare

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Antibiotic stewardship has traditionally been an inpatient concern **(Chap. 144)**, but ambulatory antibiotic use accounts for ~85% of antibiotic use by patients in most developed countries. In 2016, the Centers for Disease Control and Prevention published the "Core Elements of Outpatient Antibiotic Stewardship." The core elements include (1) committing to improving antibiotic prescribing; (2) implementing at least one policy or practice to improve antibiotic prescribing and assessing its effectiveness; (3) monitoring antibiotic prescribing and providing feedback; and (4) providing educational resources to clinicians and patients on antibiotic prescribing. Effective interventions to decrease inappropriate ambulatory antibiotic prescribing include peer comparison, accountable justification, precommitment, clinical decision support, patient education, and multifaceted interventions. Communication training has been particularly effective when it includes making a clear diagnosis, focusing on positive actions patients can take to feel better, reviewing the expected course of illness, and informing patients about concerning symptoms (red flags) for which they should seek or reconnect with care. Telemedicine—synchronous telephone or video or asynchronous electronic messaging—has the potential to improve patient convenience and reduce inappropriate antibiotic prescribing.

Several techniques that seemed promising for the reduction of ambulatory antibiotic prescribing remain unproven, have been ineffective (e.g., procalcitonin testing), or are not durable (e.g., C-reactive protein testing). The practice of delayed antibiotic prescription—i.e., a prescription given to a patient who is asked not to fill it unless symptoms do not improve in a few days—is conceptually flawed and should be avoided. Delayed antibiotic prescriptions are usually given for antibiotic-inappropriate diagnoses (e.g., viral infections); they ignore the natural history of acute respiratory infections, which are self-limited and generally last from 5 to 14 days; they put the burden of clinical decision-making on patients; and they send a confusing, mixed message to patients about the appropriateness of antibiotics for respiratory infections.

NONSPECIFIC UPPER RESPIRATORY INFECTION ("THE COMMON COLD")

DEFINITION AND ETIOLOGY

Nonspecific URI, or the common cold, is a respiratory tract infection in which no single symptom predominates. Nonspecific URI is most commonly caused by respiratory viruses that are acquired through direct contact with infected individuals, contaminated surfaces, and large and small respiratory droplets. The most common viral causes of nonspecific URIs are rhinoviruses (well over 100 serotypes; **Chap. 199**), coronaviruses, parainfluenza virus, respiratory syncytial virus, influenza virus **(Chap. 199)**, adenovirus (57 serotypes; **Chap. 199**), metapneumovirus, and bocavirus **(Chap. 199)**. Making a specific viral diagnosis is not practical, cost-effective, or necessary. Multiplex panels of reverse transcription polymerase chain reaction are available but may be overly sensitive, as prior recent infection can cause false-positive results. Although the diagnosis is usually obvious, clinicians diagnosing a nonspecific URI should also consider influenza **(Chap. 200)**, measles (cough, coryza, and conjunctivitis; **Chap. 205**), acute HIV infection (in which sore throat and rash often predominate; see below and **Chap. 202**), and COVID-19 **(Chap. 199)**.

Individual susceptibility to nonspecific URIs depends on prior exposure, immunity, general health, genetics, microbiome-related factors, and mental health and social factors, including stress. Prior exposure leads to immunity to specific rhinoviruses and adenoviruses, but the number of serotypes makes reinfection likely. Immunity to non-COVID-19 coronaviruses, parainfluenza virus, respiratory syncytial virus, and metapneumoviruses is generally weak or of short duration.

SYMPTOMS AND SIGNS

Common respiratory viruses have incubation periods of 2–8 days after exposure. Symptoms generally begin gradually and include nasal fullness or obstruction, rhinorrhea, sore throat, laryngitis, lymphadenopathy, cough, and low-grade fever. Patients may have myalgias, but this feature usually is not as prominent as it is in influenza. Epistaxis is common with frequent nose blowing.

On physical examination, findings vary, but patients may have conjunctivitis, pharyngeal erythema, pharyngeal exudates, or pharyngeal cobblestoning. Depending on the phase of illness, the nasal mucosa may be pale, boggy, or red and swollen. Nasal mucus can range from watery to purulent. On auscultation, the lungs may be clear, or the patient may have diffuse wheezing or bronchial breath sounds consistent with a viral infection. Symptoms usually last 5–10 days but often last up to 14 days.

TREATMENT

Nonspecific Upper Respiratory Infection

For adults and older children, treatment of nonspecific URI is symptom-based. Fever, myalgias, and sore throat can be treated with acetaminophen

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 35: Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat, Rachel L. Amdur; Jeffrey A. Linder
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

or pseudoephedrine. Products that combine a decongestant with analgesics, antihistamines, or both help relieve symptoms. Although supporting data are weak, cough may be relieved with dextromethorphan or benzonatate (Tessalon Perles). Opioids, while effective at relieving cough, are associated with somnolence, dysphoria, constipation, and addiction.

On physical examination, findings vary, but patients may have conjunctivitis, pharyngeal erythema, pharyngeal exudates, or pharyngeal cobblestoning.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Nonspecific Upper Respiratory Infection

For adults and older children, treatment of nonspecific URI is symptom-based. Fever, myalgias, and sore throat can be treated with acetaminophen or a nonsteroidal anti-inflammatory drug (NSAID) such as ibuprofen. Rhinorrhea can be treated with ipratropium bromide. Nasal congestion can be managed with nasal decongestants such as oxymetazoline (two sprays into each nostril twice a day for up to 5 days) or systemic decongestants such as pseudoephedrine. Products that combine a decongestant with analgesics, antihistamines, or both help relieve symptoms. Although supporting data are weak, cough may be relieved with dextromethorphan or benzonatate (Tessalon Perles). Opioids, while effective at relieving cough, are associated with somnolence, dysphoria, constipation, and addiction.

For children <6 years old, cough and cold medicines should not be prescribed, recommended, or used because of the risks of adverse effects. Honey can help soothe a sore throat for children >1 year old. Cool-mist humidifiers may help with breathing, and saline nasal drops and bulb suctioning can help with nasal congestion.

Patients need to be informed that symptoms generally peak early but can last for up to 14 days; that they are infectious as long as they have symptoms; and that they should rest and drink plenty of fluids to avoid dehydration. Red flags for which patients should seek care include a fever of >102°F, chest pain (other than from a pulled muscle), shortness of breath, dizziness, confusion, new ear or sinus pain, and symptoms lasting >14 days. Although nonspecific URI can be complicated by otitis media and bacterial sinusitis, for an individual patient, an antibiotic is more likely to cause an adverse reaction than to prevent complications.

Other remedies that are ineffective, of questionable benefit, or associated with significant adverse effects include echinacea, zinc, inhaled steam, vitamin C, vitamin D, garlic, antihistamines, Chinese medicinal herbs, intranasal glucocorticoids, *Pelargonium sidoides* herbal extract, saline nasal irrigation, and antiviral drugs.

EAR PAIN

Ear pain is most commonly caused by otitis externa and otitis media. In adults, otologic disease is almost always associated with hearing changes. At >50 years of age, temporal arteritis should be considered in patients who have headache, malaise, weight loss, fever, anorexia, and a normal ear exam. Head and neck cancers should be considered in persons with a history of smoking and alcohol use. In children, the presence of a foreign body should be considered.

Ear pain can also result from other causes of local infection, inflammation, trauma, or tumors or can be referred. Innervation of the ear and surrounding areas includes cranial nerves V, VII, IX, and X and cervical nerves C2 and C3. Neuropathic and myopathic pain syndromes (e.g., trigeminal neuralgia) can cause ear pain. Ramsay Hunt syndrome (herpes zoster oticus) **(Chap. 441)** and Bell's palsy **(Chap. 441)** are both associated with ear pain.

Dental pathology can cause pain that radiates to the ear; caries and abscesses are most common. Bruxism, malocclusion, and temporomandibular disorder may be associated with tenderness in muscular attachments and the temporomandibular joint. Salivary gland pathology and cervical adenopathy can cause pain that radiates to the ear.

Sinusitis, tonsillitis, and pharyngitis cause pain that can radiate to the ear via cranial nerve IX. Gastroesophageal reflux disease **(Chap. 321)** is often associated with ear symptoms. Myocardial infarction can cause ear pain via cranial nerve X.

Relapsing polychondritis **(Chap. 366)** is a rare condition associated with recurrent, sometimes bilateral, erythematous, or violaceous swelling of the auricle (sparing the earlobe). Inflammation from relapsing polychondritis can involve nasal septal, laryngeal, or respiratory cartilage and can cause ocular inflammation, audiovestibular damage, and nonerosive seronegative inflammatory arthritis.

OTITIS EXTERNA

Etiology and Clinical Manifestations

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Otitis externa is most common among preteen and teenage children. Risk factors for otitis externa include swimming (with the resulting condition

Otitis externa is an inflammation or infection of the external auditory canal manifesting as pain, redness, swelling, aural discharge, and hearing impairment. It is often associated with bacterial infection (frequently by *Pseudomonas aeruginosa* or *Staphylococcus aureus*), but fungi like *Aspergillus* or *Candida* can be implicated.

Otitis externa is most common among preteen and teenage children. Risk factors for otitis externa include swimming (with the resulting condition referred to as "swimmer's ear," which is more common in the summer), mechanical trauma (from cotton swabs or hearing aids), narrow ear canals, cerumen obstruction, eczema, and psoriasis. Classic swimmer's ear is associated with bacterial infection. Physical exam is notable for pain on movement of the auricle or tragus and an external auditory canal that is erythematous, edematous, inflamed, and sometimes coated with exudate on otoscopy. In contrast, fungal otitis externa often manifests with pruritus and ear discharge but without much pain.

Otitis externa can co-occur with otitis media. Preauricular, mastoid, parotid, or cervical lymphadenopathy may be present. AOM with tympanic membrane rupture (see below) can be associated with ear discharge and debris in the ear canal but (unlike otitis externa) without sensitivity to movement of the auricle.

Malignant Otitis Externa

Malignant otitis externa is a potentially life-threatening form of otitis externa that involves the temporal bone and occurs in patients with diabetes or other types of immunosuppression, often in older adults. Patients may have fever. Progression of malignant otitis externa can affect cranial nerve VII, IX, XI, or XII.

TREATMENT

Otitis Externa

Analgesia should be provided with acetaminophen or an NSAID. The mainstay of treatment is one or more topical antibacterial drugs with a glucocorticoid for 7–10 days. Polymyxin B–neomycin–hydrocortisone is often used but should be avoided in patients with tympanic membrane perforation because of ototoxicity. Ciprofloxacin–hydrocortisone is an alternative.

Topical aluminum acetate may be as effective as a topical antibacterial–glucocorticoid regimen. For patients whose condition does not improve within 2–4 days with topical treatment, ear wicks or gauze impregnated with or soaked in anti-infective agents can be placed. Ineffective treatments include oral antibiotics and topical antifungals. Otitis externa frequently recurs; its recurrence may be prevented with periodic acetic acid or aluminum acetate drops.

For malignant otitis externa, oral antipseudomonal antibiotics are often prescribed. Patients sometimes require IV pain medication, fluids, or other antimicrobials.

ACUTE OTITIS MEDIA

Epidemiology and Etiology

AOM—for which patients almost always present within days—is predominantly a disease of children, with incidence peaking at 6–24 months of age. By age 6, ~60% of children will have had an episode of AOM. Younger children appear to be susceptible because of a shorter, more horizontal eustachian tube that more easily accumulates fluid than it does in older children and adults and because their immune system is still developing.

AOM is caused by a viral URI leading to edema and inflammation of the nasopharynx and eustachian tube, collection of fluid, and infection by bacteria that colonize the nasopharynx. Viruses isolated include respiratory syncytial virus, rhinoviruses, enteroviruses, coronaviruses, influenza virus, adenoviruses, and human metapneumovirus. The bacteria most commonly isolated are *Streptococcus pneumoniae*, nontypeable *Haemophilus influenzae*, and *Moraxella catarrhalis*.

Symptoms and Signs

Symptoms of AOM include ear pain, fever, irritability, otorrhea, and anorexia. Physical examination may be notable for a bulging, inflamed, cloudy

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

least 2 days, and a temperature of >102.2°F.

Symptoms of AOM include ear pain, fever, irritability, otorrhea, and anorexia. Physical examination may be notable for a bulging, inflamed, cloudy tympanic membrane, with obscured landmarks, and immobility of the membrane on pneumatoscopy, the Valsalva maneuver, or swallowing while holding the nose shut. (An immobile tympanic membrane is also indicative of perforation, old middle-ear adhesions, a blocked auditory tube, or the presence of middle-ear fluid.) Patients have conductive hearing loss. Severe signs and symptoms include moderate to severe otalgia, otalgia lasting at least 2 days, and a temperature of >102.2°F.

AOM should be diagnosed in children with moderate to severe bulging of the tympanic membrane or new-onset otorrhea (not due to otitis externa). With mild bulging of the tympanic membrane, AOM can also be diagnosed if the patient has had symptoms for <48 h or if there is intense erythema of the tympanic membrane. AOM should *not* be diagnosed in children who do not have middle-ear effusion.

TREATMENT

Acute Otitis Media

Pain from AOM should be treated with NSAIDs or acetaminophen, which are effective for mild to moderate pain. Topical agents like benzocaine, procaine, or lidocaine may provide some additional, brief benefit beyond that offered by NSAIDs or acetaminophen.

In up to 80% of children, AOM resolves without antibiotics. Indications for antibiotic treatment in children include an age of <6 months, bilateral ear findings in children 6 months to 2 years old, otorrhea in children >6 months old, and—in children of all ages—ear findings with severe otalgia, ear pain for >48 h, or a fever of >102.2°F **(Table 35-1)**.

The benefits of antibiotics are modest and are offset by adverse effects. Antibiotics do not result in early resolution of pain but do decrease pain by day 2 or 3 (number needed to treat, 20 patients treated with antibiotics for 1 patient to have decreased pain by day 2 or 3). More children who receive antibiotics have vomiting, diarrhea, and rash (number needed to harm, 14 patients treated with antibiotics for 1 to have vomiting, diarrhea, or rash). Severe complications like mastoiditis are rare, and the number needed to treat to prevent a case of mastoiditis is ~5000 (i.e., 5000 otitis media patients treated with antibiotics to prevent 1 case of mastoiditis). The American Academy of Family Physicians recommends not routinely prescribing antibiotics for otitis media in children 2–12 years old who have nonsevere symptoms and for whom the observation option is reasonable.

The antibiotic of choice for AOM is high-dose amoxicillin (90 mg/kg per d, up to 3 g). Alternatives include cefdinir, cefuroxime, cefpodoxime, or IM ceftriaxone. If the patient has received amoxicillin in the prior 30 days, clinicians should prescribe amoxicillin/clavulanate (90/6.4 mg/kg per d) in two divided doses. The duration of antibiotic treatment is 10 days for children <2 years old or children with severe symptoms; 5–7 days for children 2–5 years old with mild to moderate AOM; and 5 days for children ≥6 years old with mild or moderate symptoms.

If a patient's condition is not better after 48–72 h of treatment, the antibiotic regimen should be changed to amoxicillin/clavulanate, a second- or third-generation oral cephalosporin, or IM ceftriaxone for 3 days. If, despite a change in antibiotics, the patient's condition still does not improve, that patient should be referred to a specialist. Middle-ear effusions are present in 60–70% of children with AOM; these should resolve over 3 months. Tympanostomy tubes should be considered for recurrent AOM (i.e., three episodes in 6 months or four episodes in 1 year). Mastoiditis is a rare complication of AOM that is suggested by postauricular tenderness, a postauricular mass, or protrusion of the ear lobe.

In adults, AOM is rare and there is little high-quality evidence to guide treatment. For adults, it remains important to differentiate AOM from OME, but AOM is generally treated with antibiotics, regardless of bilaterality or otorrhea. Amoxicillin is the drug of choice. Adults should also be treated with decongestants and analgesics. Adults with more than two episodes in a year or persistent effusion should be referred to an otolaryngologist.

TABLE 35-1

Indications for Antibiotic Treatment of Acute Otitis Media

AGE	INDICATION
<6 months	Antibiotic treatment reasonable for all
>2 years	Symptoms worsening or not improving within 48–72 h

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 35-1

Indications for Antibiotic Treatment of Acute Otitis Media

AGE	INDICATION
<6 months	Antibiotic treatment reasonable for all
6 months to 2 years	Bilateral ear findings
≥6 months	Otorrhea
>2 years	Symptoms worsening or not improving within 48–72 h
All ages	Ear findings with severe otalgia, otalgia lasting at least 2 days, or temperature of >102.2°F

OTITIS MEDIA WITH EFFUSION

Definition and Etiology

OME, also called serous otitis media, occurs when there is fluid in the middle ear but no acute infection. Most patients with OME are young children; >60% of cases occur in children <2 years old. Many children have recurrent episodes.

OME is most often a sequela of a viral infection causing AOM, but it can also be caused by allergies. In addition to allergies, predisposing factors include craniofacial abnormalities, gastroesophageal reflux, and enlarged adenoids.

Symptoms and Signs

The most common symptoms are decreases in sound conduction and hearing. Children with OME may exhibit impaired language development or communication difficulties. More rarely, patients complain of intermittent ear fullness or earache, tinnitus, or balance problems. On examination, the tympanic membrane may be translucent or gray with fluid (often colorless or amber), air-fluid levels, or bubbles behind the membrane. There is a loss of the light reflex. The tympanic membrane has decreased mobility on pneumatic otoscopy. The evaluation may include audiometry, tympanometry, and, in infants, measurement of auditory brainstem responses.

OME usually resolves spontaneously within 4–6 weeks. If it persists for >3 months, the condition is referred to as chronic OME or chronic serous otitis media.

Cholesteatomas are accumulations of epithelium or keratin in the middle ear that can enlarge, perforate the tympanic membrane, envelop the ossicles, or destroy surrounding tissue. Cholesteatomas can cause labyrinthitis, hearing loss, cranial nerve palsies, vertigo, meningitis, extradural or brain abscess, and lateral sinus thrombophlebitis.

TREATMENT

Otitis Media with Effusion

OME is treated with myringotomy with tympanostomy tube insertion. For young children with nasal obstruction or recurrent infection, adenoidectomy may be considered. Medications, including antihistamines, glucocorticoids, or antibiotics, do not reliably help. Children at risk for speech or language delay may need earlier referral for more aggressive treatment.

ACUTE MASTOIDITIS

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

children <2 years old but can occur at any age. Acute mastoiditis is often a complication of AOM but may develop without clinically apparent, prior AOM. In older children with acute mastoiditis, clinicians should suspect cholesteatoma.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Etiology

Acute mastoiditis is a serious infection with significant morbidity despite antibiotic and surgical treatment. This condition is most common among children <2 years old but can occur at any age. Acute mastoiditis is often a complication of AOM but may develop without clinically apparent, prior AOM. In older children with acute mastoiditis, clinicians should suspect cholesteatoma.

The pathogenesis of mastoiditis involves spread of organisms from the middle-ear spaces through the aditus ad antrum to the mastoid air cells. *Incipient* mastoiditis consists of fluid within the mastoid air cells, without bony destruction of the bony septa, and can progress to *coalescent* mastoiditis, with destruction of the bony septa. Acute mastoiditis often causes subperiosteal abscess laterally. The organisms most commonly involved in mastoiditis are *S. pneumoniae*, *Streptococcus pyogenes, H. influenzae, S. aureus* (including methicillin-resistant *S. aureus* [MRSA] strains), and *P. aeruginosa*.

Symptoms and Signs

Symptoms of acute mastoiditis include ear pain, fever, lethargy, or fussiness despite adequate treatment of AOM. Patients—especially those with subperiosteal abscess—may have postauricular erythema, tenderness, warmth, fluctuance, and protrusion of the auricle. Otoscopic examination most often yields findings of AOM and may show superoposterior protrusion of the external auditory canal. Complications of mastoiditis include facial nerve palsy, labyrinthitis, skull osteomyelitis, temporal lobe abscess, cerebellar abscess, meningitis, epidural abscess, subdural abscess, venous sinus thrombosis, or Bezold's abscess (an abscess medial to the sternocleidomastoid that tracks into the deep cervical fascia).

Evaluation

Laboratory evaluation reveals elevation of inflammatory markers and white blood cells with neutrophilia. Imaging is not necessary in children with a classic history and presentation but may be required if there is concern about complications or severity. CT may show disruption of bony septations, fluid, mucosal thickening, periosteal thickening, disruption of the periosteum, or subperiosteal abscess. MRI with gadolinium permits better visualization of abscesses and vascular problems.

Differential Diagnosis

The differential diagnosis of acute mastoiditis includes cellulitis, otitis externa, postauricular lymphadenopathy, perichondritis, and tumors, including rhabdomyosarcoma, Ewing sarcoma, and myofibroblastic tumor.

TREATMENT

Mastoiditis

Patients with mastoiditis should be admitted to the hospital and treated with IV antibiotics and myringotomy, with or without tympanostomy tubes; if there is no improvement within 48 h, mastoidectomy should be undertaken. Tympanostomy or myringotomy samples or subperiosteal abscess drainage should be sent for culture and sensitivity testing. Depending on complications, additional drainage and surgical procedures may be necessary.

Empirical IV antibiotic therapy for children without recurrent AOM or recent antibiotic treatment consists of vancomycin (if there is concern about antibiotic-resistant *S. pneumoniae* or MRSA) or a cephalosporin (e.g., cefepime or ceftazidime). Patients with recurrent AOM or recent antibiotic treatment should be given vancomycin plus an antipseudomonal penicillin. Culture and sensitivity results will guide antibiotic changes. IV antibiotic therapy should be continued for 7–10 days, and patients should complete a 4-week course of oral antibiotics.

SINUS SYMPTOMS

Sinus symptoms are commonly due to respiratory viruses. These symptoms are considered acute if they last <4 weeks, subacute if they last 4–12 weeks, and chronic if they last ≥12 weeks. Beyond sinus infection, the differential diagnosis of rhinitis includes the common cold, allergic rhinitis (Chap. 352), vasomotor rhinitis, rhinitis medicamentosa due to topical decongestants, drug-induced rhinitis (e.g., due to aspirin, ibuprofen, or beta

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 35: Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat, Rachel L. Amdur; Jeffrey A. Linder
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

referral of symptoms to the sinuses. Patients who have uncontrolled diabetes or are otherwise immunocompromised can have rapidly progressing invasive fungal infections (Chap. 211). More indolent fungal infections should be considered in the event of recurrent or nonresolving sinusitis. In children, it is important to consider the presence of a foreign body as a cause of sinus symptoms.

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

weeks, and chronic if they last ≥12 weeks. Beyond sinus infection, the differential diagnosis of rhinitis includes the common cold, allergic rhinitis (Chap. 352), vasomotor rhinitis, rhinitis medicamentosa due to topical decongestants, drug-induced rhinitis (e.g., due to aspirin, ibuprofen, or beta blockers), autoimmune disease (e.g., granulomatosis with polyangiitis), and cerebrospinal fluid leak. Pain over the sinuses can be caused by headaches (Chap. 430), facial pain syndromes, temporomandibular disorder (Chap. 36), and dental pathology. Gastroesophageal reflux can cause referral of symptoms to the sinuses. Patients who have uncontrolled diabetes or are otherwise immunocompromised can have rapidly progressing invasive fungal infections (Chap. 211). More indolent fungal infections should be considered in the event of recurrent or nonresolving sinusitis. In children, it is important to consider the presence of a foreign body as a cause of sinus symptoms.

ACUTE SINUSITIS

Definition and Etiology

Sinusitis is an inflammation of the paranasal sinuses; *rhinosinusitis* also involves the nasal passages. The majority of acute sinusitis cases are caused by respiratory viruses. A diagnosis of sinusitis is a major reason for unnecessary antibiotic prescribing in adults: although <2% of sinusitis episodes are due to bacteria (most often *S. pneumoniae, H. influenzae,* or *M. catarrhalis*), antibiotics are prescribed at >70% of office visits for sinusitis. According to guideline criteria, no more than 50% of adults—and probably closer to 20%—meet the criteria for antibiotic prescribing.

Symptoms and Signs

Sinusitis symptoms commonly include purulent nasal discharge, facial congestion or fullness, and facial pain or pressure. Other symptoms include fever; hyposmia or anosmia; ear pain, pressure, or fullness; postnasal drip; halitosis; maxillary toothache; cough; and fatigue. Risk factors for developing sinusitis include an age of 45–65 years, smoking, asthma, air travel, and allergies.

On physical examination, direct rhinoscopy reveals excess mucus or purulence. Patients may have tenderness over the maxillary sinuses and, in severe cases, erythema and swelling of the maxilla. Sinus transillumination is not accurate in diagnosing sinusitis.

Complications

Complications from sinusitis can be dramatic but are extremely rare. These complications may include orbital cellulitis, osteomyelitis, meningitis, intracranial abscesses, and cavernous sinus thrombosis. New symptoms that might indicate a sinusitis complication include confusion, unilateral weakness, proptosis, limited ocular movements, and acute vision changes.

RECURRENT ACUTE SINUSITIS

Patients who have four or more episodes of acute sinusitis in a year, without signs or symptoms between episodes, are said to have recurrent acute sinusitis.

INVASIVE FUNGAL SINUSITIS

Invasive fungal sinusitis may develop in immunocompromised patients, such as those with uncontrolled diabetes or transplant recipients, and should be considered an emergency. Invasive fungal sinusitis is caused by Mucorales fungi or *Aspergillus* (Chap. 217). Patients may appear to have a rapidly progressive case of rhinosinusitis, with facial pain and pressure, headaches, and fever followed within days by cranial nerve involvement, orbital swelling, cellulitis, proptosis, chemosis, and ophthalmoplegia. Patients may be critically ill. Evaluation should include nasal endoscopy with biopsy and imaging with gadolinium-enhanced MRI as the preferred modality.

NOSOCOMIAL SINUSITIS

Nosocomial sinusitis occurs in critically ill patients, often those who are nasotracheally intubated. Nosocomial sinusitis should be suspected in hospitalized patients who have fever without another identifiable cause.

TREATMENT

Acute Sinusitis

All patients with acute sinusitis should be counseled about symptom-based treatments, which may include decongestants, analgesics/antipyretics like acetaminophen or NSAIDs (e.g., ibuprofen), nasal saline spray, and nasal washes provide relief. Intranasal glucocorticoids may help, particularly for patients with an allergic cause of sinusitis. Because patients may be accustomed to receiving antibiotics, provision of a clear

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 35: Upper Respiratory Symptoms, Including Earache, Sinus Symptoms, and Sore Throat, Rachel L. Amdur; Jeffrey A. Linder
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Acute Sinusitis

All patients with acute sinusitis should be counseled about symptom-based treatments, which may include decongestants, analgesic/antipyretics, nasal saline, or intranasal glucocorticoids. Intranasal decongestants (e.g., oxymetazoline, two sprays in each nostril twice a day for no more than 5 days) and oral decongestants (e.g., 12-h pseudoephedrine [120 mg] during the day) relieve pain, pressure, and rhinorrhea. Analgesics and antipyretics like acetaminophen or NSAIDs (e.g., ibuprofen), nasal saline spray, and nasal washes provide relief. Intranasal glucocorticoids may help, particularly for patients with an allergic cause of sinusitis. Because patients may be accustomed to receiving antibiotics, provision of a clear explanation, symptom-based treatments, and reasons for reconsultation are important. Red flags for which patients should reconsult include recurrent fever of >102°F, sinus symptoms that worsen after initial improvement, and rapid worsening of facial pain that becomes persistent, as well as any other concerning symptoms.

Antibiotic prescribing criteria for sinusitis are based on symptoms **(Table 35-2)**. Only patients with persistent, severe, or worsening symptoms, especially those who have already used decongestants and analgesics for 2–4 days, meet the criteria for antibiotic prescribing. The antibiotic of choice is amoxicillin/clavulanate (875 mg/125 mg bid for 7 days). Amoxicillin (875 mg PO bid for 7 days) is an alternative. For patients with mild penicillin allergies, cefuroxime is a reasonable choice. For those with severe penicillin allergies, doxycycline is a reasonable alternative. Macrolides are specifically not recommended for sinusitis because of high rates of macrolide-resistant *S. pneumoniae.*

Patients who meet the criteria for antibiotic prescribing should show signs of improvement after 3–5 days of therapy. If not, second-line regimens include amoxicillin/clavulanate (2000 mg/125 mg bid for 7 days) or levofloxacin, although fluoroquinolones are associated with dysglycemia, neuropathy, and tendon and aortic rupture. For patients whose condition still is not improving after 3–5 days of treatment with a second-line antibiotic or in whom a complication or an alternative diagnosis is suspected, clinicians should consider referral to an otorhinolaryngologist and/or the performance of imaging tests. The imaging modality of choice is noncontrast CT. Patients with recurrent acute sinusitis may benefit from nasal culture during episodes; imaging between episodes to identify predisposing anatomic abnormalities; and allergic or immunologic evaluation.

Patients with acute fungal sinusitis should be treated with IV antifungal agents and often require surgical debridement. Patients with nosocomial sinusitis should have precipitating factors (e.g., nasotracheal intubation) addressed and should be empirically treated with broad-spectrum antibiotics until culture and susceptibility results are available.

TABLE 35-2

Indications for Antibiotic Treatment of Acute Sinusitis

INDICATION	DEFINITION
Persistent	Symptoms lasting ≥10 days
Severe	Fever of >102°F and either purulent nasal discharge or nasal pain for at least 3–4 consecutive days
Worsening	New fever, headache, or increase in nasal discharge following an upper respiratory tract infection that lasted for 5–6 days and was initially improving

Note: In typical populations, roughly 20% and no more than 50% of adults with sinusitis will meet the criteria for antibiotic prescribing.

CHRONIC SINUSITIS

Definition and Etiology

Chronic sinusitis is defined as inflammation of the paranasal sinuses that lasts >12 weeks. Chronic sinusitis is primarily an inflammatory disease and

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

occupational factors, and other airborne toxins. Functional or immunologic problems can include impaired mucociliary clearance (e.g., due to cystic fibrosis) or immunodeficiency due to acquired conditions or medications. Chronic sinusitis often coexists with allergic rhinitis and asthma.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chronic sinusitis is defined as inflammation of the paranasal sinuses that lasts >12 weeks. Chronic sinusitis is primarily an inflammatory disease and can also be associated with acute or chronic infection or allergic, structural (e.g., deviated nasal septum or polyps), and immunologic etiologies. Repeated viral infections may lead to chronic sinusitis. Bacterial colonization or chronic infection plays a role in some cases of chronic sinusitis. *S. aureus* and gram-negative bacteria are commonly identified. Commonly involved allergens and irritants are dust mites, mold, tobacco smoke, occupational factors, and other airborne toxins. Functional or immunologic problems can include impaired mucociliary clearance (e.g., due to cystic fibrosis) or immunodeficiency due to acquired conditions or medications. Chronic sinusitis often coexists with allergic rhinitis and asthma.

Symptoms and Signs

Cardinal symptoms of chronic sinusitis are facial pain or pressure, nasal discharge or postnasal drip, congestion, and hyposmia or anosmia. Associated symptoms may include fatigue, malaise, ear pressure, hoarseness, and cough. The diagnosis of sinus inflammation must be confirmed with anterior rhinoscopy, nasal endoscopy, or imaging because up to 40% of patients with chronic sinus symptoms do not have mucosal changes evidencing disease.

In practical terms, chronic sinusitis can be divided into three main types (in decreasing order of frequency): (1) chronic sinusitis without polyps, (2) chronic sinusitis with polyps, and (3) allergic fungal sinusitis. In general, chronic sinusitis without polyps is more common among women, develops in childhood and young adulthood, is characterized by presentations with facial pain, and is often due to T_H1 lymphocyte predominance associated with bacterial infection or colonization. Chronic sinusitis with polyps is more common among men; develops in adulthood; is characterized by presentations with decrease or loss of smell, asthma, or aspirin sensitivity **(Chap. 287)**; and is often due to T_H2 lymphocyte predominance associated with eosinophilic inflammation, asthma, or aspirin sensitivity. Allergic fungal rhinosinusitis is also associated with polyp formation; typically occurs in patients in their 20s and 30s who are from warm, humid regions and who have other atopic diseases; and is associated with IgE-mediated allergy and eosinophils **(Chap. 217)**. The mucus in allergic fungal rhinosinusitis is classically greenish-brown, has a peanut butter–like consistency, and includes viable hyphae from *Aspergillus* or other fungal species. Allergic fungal rhinosinusitis is resistant to medical treatments.

Evaluation

On anterior rhinoscopy, polyps are seen as white, gray, tan, or yellow translucent growths in the middle meatus. The imaging modality of choice is noncontrast CT. Allergic fungal rhinosinusitis may be unilateral; however, unilateral symptoms or polyps on exam or imaging, especially if associated with bloody discharge, should raise concern about tumors.

TREATMENT

Chronic Sinusitis

Treatment includes avoidance of identifiable triggers such as allergens, smoke, and irritants. Saline sprays and washes provide symptom relief, and higher-volume saline washes are probably more effective. Intranasal glucocorticoids, including mometasone and fluticasone sprays or higher-potency and higher-volume budesonide rinses, are mainstays of treatment, especially for chronic sinusitis with polyps. Intranasal glucocorticoids reduce polyp size. Oral administration of glucocorticoids for 2–3 weeks is sometimes effective against chronic sinusitis that is unresponsive to intranasal steroids—again, especially for patients with polyps. Intranasal or systemic antihistamines may help patients whose illness has an allergic component. Likewise, leukotriene antagonists like montelukast may help.

Although antibiotics are frequently prescribed for 2–4 weeks to patients with chronic sinusitis, there is little evidence that these drugs are effective. Evidence of modest quality supports the use of 3 months of macrolide treatment for patients who have chronic sinusitis without polyps. Antifungal agents have not shown benefit against any subtype of chronic sinusitis. Decongestants should be used only sparingly and briefly.

Endoscopic sinus surgery improves quality of life in patients who have had inadequate responses to medical therapy. Patients with more limited, focal disease may more reliably have better results. The goals of surgery are to remove polyps from the nasal cavity and paranasal sinuses. For patients with allergic fungal rhinosinusitis, medical therapy is classically ineffective, surgery produces good results, and patients should be treated with perioperative glucocorticoids. In children, adenoidectomy may be effective in some cases. In the future, immune endotyping may allow selection of more individualized biological treatments.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Sore throat is not synonymous with pharyngitis and can also be caused by submandibular space, retropharyngeal and peritonsillar abscesses, thyroiditis, gastroesophageal reflux, tumors, and postnasal drainage.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

SORE THROAT AND NECK PAIN

Sore throat is not synonymous with pharyngitis and can also be caused by submandibular space, retropharyngeal and peritonsillar abscesses, thyroiditis, gastroesophageal reflux, tumors, and postnasal drainage.

Acute pharyngitis, in which symptoms are generally present for days, is most often caused by respiratory viruses; is often caused by group A β-hemolytic streptococci (GAS); and can be caused by other bacteria (including *Neisseria gonorrhoeae*), Epstein-Barr virus (EBV), and HIV. On physical examination, pharyngeal erythema is associated most commonly with viral infections, including the common cold and influenza. Pharyngeal exudate should not be confused with *Candida* infection, which looks like cottage-cheese, can be scraped off, and leaves a bleeding surface, or leukoplakia, which cannot be scraped off. History and exam findings may help differentiate sore throat and pharyngitis of various etiologies **(Table 35-3)**.

TABLE 35-3

Clinical Findings That Suggest Various Forms of Nonstreptococcal Pharyngitis

CLINICAL FINDING(S) OR BEHAVIORAL FACTOR	SUSPECTED DIAGNOSIS
Scarlatiniform rash	Group A β-hemolytic streptococci or *Arcanobacterium haemolyticum*
Cough and otitis media	*Haemophilus influenzae*
Sex between men with associated urogenital symptoms, fellatio between a woman and a man who has current urogenital symptoms, persistent sore throat unresponsive to penicillin	*Neisseria gonorrhoeae*
Travel to endemic areas, pseudomembrane on examination	*Corynebacterium diphtheriae*
Persistent sore throat with bronchopulmonary symptoms	*Mycoplasma pneumoniae*
Marked adenopathy (especially that involving posterior cervical or auricular nodes), splenomegaly, palatine petechiae, gelatinous uvula	Acute infectious mononucleosis
New sexual partner in the previous month; fever, rash, myalgias, headache	Acute HIV infection

STREPTOCOCCAL PHARYNGITIS

GAS is the only common cause of sore throat that should be treated with antibiotics. The principal goal in the evaluation of adults with sore throat is to identify patients likely to have GAS pharyngitis, or "strep throat." Prompt antibiotic treatment of adults likely to have strep throat has the potential to reduce symptoms, prevent the spread of disease, and reduce suppurative complications (e.g., peritonsillar abscess). Nonsuppurative complications are rare. In developed countries, the prevalence of rheumatic fever **(Chap. 148)** is extremely low, and antibiotic treatment does not prevent poststreptococcal glomerulonephritis **(Chap. 148)**.

Most patients with non-GAS pharyngitis have various forms of viral pharyngitis and do not require antibiotics. Nevertheless, clinicians prescribe antibiotics to a majority of adults with sore throats. By using a simple clinical scoring algorithm, clinicians can predict the presence or absence of GAS with sufficient accuracy and avoid prescribing antibiotics to patients who are unlikely to have strep throat. Although there is a role for testing (see "Evaluation," below), most adults with sore throat do not need to have a GAS test.

About 10% of adults with sore throat are infected with GAS. Among children with sore throat, the prevalence of GAS can be as high as 35%, with rates peaking from 5 to 15 years of age. The prevalence of GAS is higher in winter and early spring. The risk of streptococcal pharyngitis is elevated among health care and child care workers, teachers, parents of young children, and patients exposed to individuals with strep throat. Clinicians need to be

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Evaluation

peaking from 5 to 15 years of age. The prevalence of GAS is higher in winter and early spring. The risk of streptococcal pharyngitis is elevated among health care and child care workers, teachers, parents of young children, and patients exposed to individuals with strep throat. Clinicians need to be aware of local outbreaks of GAS infection, particularly in military and institutional settings, where the prevalence of GAS and the risk of acute rheumatic fever may be elevated.

Evaluation

The Centor criteria consist of four findings, each of which is assigned 1 point: (1) history of fever, (2) absence of cough, (3) tender anterior cervical lymphadenopathy, and (4) tonsillar exudate or swelling. The Centor criteria are easy to assess and accurately stratify adult patients with suspected streptococcal pharyngitis. Patients with no points have a 2% probability of being infected with GAS, whereas those with 4 points have a probability of 41% **(Table 35-4)**. The Centor criteria have an area under the curve of 0.79. Other clinical decision algorithms similar to the Centor criteria may not perform as well, are not as simple, or have not been as rigorously evaluated.

TABLE 35-4

The Centor Criteria and the Probability of Streptococcal Pharyngitis for Adults[a]

NO. OF CRITERIA MET[b]	POSTEVALUATION PROBABILITY (%)	RECOMMENDATION
0	2	No test, no antibiotic
1	3	No test, no antibiotic
2	8	Rapid test
3	19	Rapid test
4	41	Empirical antibiotic treatment or rapid test

[a]Assuming a pretest probability of strep throat for adults of 10%. [b]The criteria are (1) a history of fever, (2) an absence of cough, (3) tender anterior cervical lymphadenopathy, and (4) tonsillar swelling or exudate. Each criterion gets 1 point. Roughly 40–60% of adults will meet no criteria or one criterion; ~20% will meet the criteria for antibiotic prescribing.

If the test/no treatment threshold is set at 5%, for a GAS prevalence of ~10%, adults meeting no criteria or only one Centor criterion have a probability of GAS pharyngitis so low that they should neither be tested nor be treated with an antibiotic. Adults meeting two or three Centor criteria have an intermediate probability of GAS pharyngitis; they should have a rapid antigen test performed, and the results should guide antibiotic treatment. For adults meeting four Centor criteria, it is reasonable either to perform a rapid test or to institute empirical antibiotic treatment. However, some guidelines recommend—and some ambulatory quality measures require—a GAS test to be associated with antibiotic prescribing in adults, regardless of the number of Centor criteria met.

In children, the Centor criteria are less specific, and streptococcal pharyngitis should be confirmed with testing. Children who have signs of pharyngitis without signs of viral infection (conjunctivitis, runny nose, cough, hoarseness, nonexudative oral lesions) should have testing performed.

Outside of the United States, because complications are rare and even streptococcal pharyngitis is self-limited in the vast majority of cases, some guidelines do not recommend use of rapid GAS testing or routine antibiotic treatment of sore throat.

Clinicians should have a lower threshold for diagnosing and treating GAS pharyngitis in patients with a history of acute rheumatic fever, patients with documented streptococcal exposure in the past week, patients who live in a community with a current strep throat epidemic, and patients who are diabetic or otherwise immunocompromised.

RAPID STREP TESTS

Rapid GAS-specific antigen tests have a sensitivity of ~80% and a specificity of ~95%. Results are available within minutes and can be used to make

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

THROAT CULTURES

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Rapid GAS-specific antigen tests have a sensitivity of ~80% and a specificity of ~95%. Results are available within minutes and can be used to make therapeutic decisions before the patient leaves the office. Improper collection technique can adversely affect the sensitivity of rapid strep tests: clinicians should rub the tonsils and pharynx, touching any areas where exudate or ulceration are present.

THROAT CULTURES

A single-swab throat culture has a sensitivity of ~85–90%, as defined by isolation of GAS on a second swab. A throat culture can also be falsely positive for true infection: some patients with a culture positive for GAS may be only uninfected carriers, as defined by their failure to exhibit a fourfold increase in antibodies to GAS—the gold standard test. Among adults and children seeking medical care for a sore throat, test specificity may be as low as 50–70% because of patients who do not exhibit serologic evidence of infection. Throat cultures are not recommended for the routine evaluation of adults with sore throat. The modest gain in sensitivity over rapid testing is outweighed by the 24- to 48-h delay in test results, with a consequent delay in the symptomatic relief associated with antibiotic treatment.

Indiscriminate strep testing in adults with sore throat or respiratory symptoms should be discouraged. Rapid strep tests and culture do not differentiate between patients who have true infection and those who are carriers of GAS (with carriage rates as high as 20% among schoolchildren and ~5% among adolescents and young adults). In adults who meet no Centor criteria or only one criterion—40–60% of adults with pharyngitis—a positive test is highly likely to be falsely positive and/or to represent GAS carriage.

Complications

Complications of streptococcal pharyngitis are rare but include acute rheumatic fever **(Chap. 148)**, poststreptococcal glomerulonephritis **(Chap. 148)**, scarlet fever **(Chap. 148)**, sinusitis, peritonsillar abscess, and other invasive GAS infections.

TREATMENT

Streptococcal Pharyngitis

All patients with pharyngitis—nonstreptococcal and streptococcal—should receive analgesics (acetaminophen or NSAIDs). Saline gargles, humidification, soft foods, and tea with honey soothe a painful throat.

Penicillin is the antibiotic of choice for streptococcal pharyngitis **(Table 35-5)**. Penicillin is a narrow-spectrum, low-cost, and well-tolerated drug to which no GAS isolate has been resistant. Amoxicillin is an acceptable alternative in children as it comes in a palatable liquid form. For patients with mild penicillin allergy, cephalexin and cefadroxil are good alternatives. For patients with severe penicillin allergies, clinicians should prescribe erythromycin, clarithromycin, or clindamycin. Unlike other infections for which emerging evidence supports progressively shorter antibiotic courses, streptococcal pharyngitis requires longer courses (7–10 days), which are more effective.

Glucocorticoids (e.g., dexamethasone, 10 mg as a single oral dose) have so far been poorly studied as an adjunctive treatment for sore throat and strep throat and are not recommended. These drugs may result in decreased pain within 24 h but do not decrease school or work absenteeism or relapse rates. Even short courses of steroids are associated with increased rates of sepsis, gastrointestinal bleeding, congestive heart failure, venous thromboembolism, and fracture within 30 days.

Streptococcal and nonstreptococcal pharyngitis should resolve in 3–5 days. Symptoms that should lead patients to seek further care include shaking chills (rigors), neck swelling (beyond lymphadenopathy), trouble swallowing, drooling, or symptoms that persist for >5 days without improvement.

TABLE 35-5

Antibiotic Treatment of Group A Streptococcal Pharyngitis

ANTIBIOTIC	DOSING
Antibiotic of Choice	
Alternative for Non-Penicillin-Allergic Patients	

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 35-5

Antibiotic Treatment of Group A Streptococcal Pharyngitis

ANTIBIOTIC	DOSING
Antibiotic of Choice	
Penicillin	500 mg PO qid or 1000 mg PO bid × 10 days
Alternative for Non-Penicillin-Allergic Patients	
Amoxicillin	500 mg PO bid or 1000 mg qd × 10 days
Alternatives for Non-Anaphylactic Penicillin-Allergic Patients	
Cephalexin	500 mg PO bid × 10 days
Cefadroxil	1 g PO qd × 10 days
Alternatives for Patients with Severe Penicillin Allergy	
Erythromycin	250–500 mg PO qid or 500–1000 mg PO bid × 5 days
Clarithromycin	500 mg PO bid × 5 days
Clindamycin	300 mg PO tid × 10 days

NONSTREPTOCOCCAL PHARYNGITIS

Acute Infectious Mononucleosis

New EBV infection may be the cause of pharyngitis in 1–6% of young adults **(Chap. 194)**. EBV is rarely the cause of pharyngitis in adults >40 years of age. The full-blown acute syndrome, which is present in only about one-fourth of patients with infectious mononucleosis (“mono”), is characterized by a triad of clinical, hematologic, and serologic findings. The clinical presentation is typified by the development over several days of malaise, fever, sore throat, and marked adenopathy that is particularly evident in the cervical lymph nodes. On physical examination, marked adenopathy is virtually always documented and is most specific for mononucleosis when the posterior cervical or posterior auricular nodes are involved. Splenomegaly and exudative pharyngitis with prominent tonsillar swelling, palatine petechiae, and a gelatinous uvula are often noted. The classic hematologic findings are an absolute lymphocyte count of >4000/μL or a relative lymphocyte count of >50% with “atypical” morphologic features in >10% of the lymphocytes. The characteristic serologic finding is the heterophil antibody, which is detectable in only 40% of patients during the first week of illness but in 80–90% of patients by the third week.

Other Bacterial Pharyngitis

Non–group A streptococci (especially group C and group G streptococci), *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, *N. gonorrhoeae*, and *H. influenzae* have all been associated with sore throat in some studies. Although antibacterial treatment has not been proven to speed the resolution of symptoms and signs of any of these types of nonstreptococcal pharyngitis, antibiotic treatment is indicated if throat cultures from a patient with persistent sore throat yield group C or group G streptococci.

LEMIERRE’S SYNDROME

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

syndrome probably occurs predominantly in male patients. Clinicians should consider Lemierre’s syndrome in a teenage or young adult patient who has non-GAS pharyngitis that is not resolving, particularly if it is accompanied by rigors, neck pain or swelling, or other extrapharyngeal symptoms.

Lemierre's syndrome consists of septic thrombophlebitis of the internal jugular vein accompanied by metastatic infections, most commonly of the lung but with possible involvement of the joints, bones, liver, meninges, and brain. Lemierre's syndrome is most commonly caused by *Fusobacterium necrophorum*, although it can also be caused by species of *Bacteroides, Eikenella, Streptococcus, Peptostreptococcus,* or other bacterial genera. This syndrome probably occurs predominantly in male patients. Clinicians should consider Lemierre's syndrome in a teenage or young adult patient who has non-GAS pharyngitis that is not resolving, particularly if it is accompanied by rigors, neck pain or swelling, or other extrapharyngeal symptoms.

GONOCOCCAL PHARYNGITIS

N. gonorrhoeae may be the cause of pharyngitis in 1% of adult patients seeking primary care for a sore throat, although gonococcal infection of the pharynx is more often asymptomatic. When symptomatic, pharyngeal gonorrhea may range from mild to severe, with protracted pharyngitis characterized by pain, fever, and pharyngeal exudate. Gonococcal pharyngitis should be suspected in men who have sex with men with associated symptoms of urogenital infection, women who have practiced fellatio with a man with genital gonorrhea, and anyone who has persistent sore throat that has been unresponsive to treatment for presumptive streptococcal pharyngitis.

DIPHTHERIA

Diphtheria, caused by *Corynebacterium diphtheriae,* is endemic in developing countries **(Chap. 150)**. Diphtheria produces only mild pharyngitis beneath its characteristic grayish pseudomembrane.

ACUTE HIV INFECTION

Clinicians should consider acute HIV infection in patients with sore throat, particularly when it is associated with headache, fever, myalgias, lymphadenopathy, anorexia, and rash **(Chap. 202)**. Of patients with acute HIV infection, roughly half have a sore throat. However, in most settings in the United States, only ~1% of patients with viral or mononucleosis-like symptoms have acute HIV infection.

HEAD AND NECK ABSCESSES

Head and neck abscesses are more common among patients with diabetes, who are immunocompromised, and among older adults. Such abscesses are often a complication of infections of the teeth and gums, throat, or salivary ducts; lymphadenitis; ear infections; sinus infections; congenital cysts; and IV drug use. Prompt recognition is important, as head and neck abscesses can cause airway compromise due to edema or mass effect. Head and neck abscesses can follow fascial planes and spread to the mediastinum (where they can cause mediastinitis, pleural effusions, empyema, or pericarditis), the carotid sheath, the skull base, and the meninges. Head and neck abscesses have also been associated with aspiration pneumonia, necrotizing fasciitis, Lemierre's syndrome, and toxic shock syndrome.

Submandibular abscesses generally result from an infected or extracted tooth and can cause Ludwig angina, a swelling of the floor of the mouth that can enlarge and displace the tongue posteriorly.

Peritonsillar abscesses, which may occur predominantly in male patients, generally result from complicated bacterial pharyngitis and present with fever, dysphagia, profound throat pain (necessitating drooling to avoid swallowing saliva), trismus, and "hot potato voice" (inability to articulate, as if patients have hot food in their mouths). Patients are likely to have unilateral palate bulging, often with uvular deviation. Peritonsillar abscesses are caused by viridans group streptococci, β-hemolytic streptococci, *F. necrophorum*, *S. aureus, Prevotella,* and *Bacteroides.*

Retropharyngeal abscesses often present after an antecedent URI in children with sore throat, dysphagia, deep neck pain, neck stiffness, trismus, and drooling. The pharyngeal wall may be displaced, but swelling or abscess may not be apparent on examination. In severe cases, patients may have dyspnea and stridor.

Patients with suspected head and neck abscesses, with the possible exception of patients who have obvious peritonsillar abscesses, should undergo imaging by CT.

TREATMENT

Head and Neck Abscesses

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

IV antibiotics should cover streptococci, anaerobes, and possibly *S. aureus.* Frequently used antibiotics include ampicillin/sulbactam, clindamycin plus ceftriaxone, or meropenem. For some abscesses with adequate source control with incision and drainage, penicillin may be as effective as

Access Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Head and Neck Abscesses

The mainstays of treatment for head and neck abscesses are securing the airway, surgical drainage, and IV antibiotic administration. To secure the airway, mask ventilation or oral intubation may not be effective, and oral fiberoptic intubation or tracheotomy may be necessary. Peritonsillar abscess may be managed with needle aspiration and/or tonsillectomy. Other head and neck abscesses require incision and drainage. The selected IV antibiotics should cover streptococci, anaerobes, and possibly *S. aureus*. Frequently used antibiotics include ampicillin/sulbactam, clindamycin plus ceftriaxone, or meropenem. For some abscesses with adequate source control with incision and drainage, penicillin may be as effective as broader-spectrum agents.

EPIGLOTTITIS

Along with associated dysphagia, odynophagia, hoarseness, and stridor or tachypnea, supraglottitis or epiglottitis must be considered in adults presenting with sore throat. The inflamed and enlarged epiglottis protrudes up into the oropharynx. Patients may extend their neck or lean forward and drool oral secretions to avoid swallowing. Epiglottitis can cause "hot potato voice." Attempts to examine or swab the posterior pharynx or obtain a culture can provoke laryngospasm and should only be done carefully in a controlled setting. Because obstruction of the airway may become acutely life-threatening, the patient with epiglottitis must be observed in a hospital setting, and examination in an operating room, where an airway can be established immediately by an experienced operator, should be strongly considered. Although not necessary for the diagnosis, a lateral neck radiograph can demonstrate epiglottal swelling referred to as the "thumb sign."

In adults, conservative therapy under observation is sufficient in most cases, but intubation by an experienced clinician or tracheostomy may become necessary. Treatments also include humidification with nebulized normal saline or humidified oxygen and administration of glucocorticoids, IV antibiotics, and nebulized epinephrine.

H. influenzae, the most common cause of supraglottitis in children, is less common in adults. Other responsible organisms in adults are *S. pneumoniae*, *S. pyogenes*, and *S. aureus*. The *H. influenzae* type b vaccine has led to a dramatic decrease in epiglottitis overall, with large reductions in young children; however, the incidence of supraglottitis and epiglottitis in adults may be increasing.

LARYNGITIS

Laryngitis—inflammation of the larynx and surrounding structures—is most commonly caused by viral URIs. In children, parainfluenza virus can cause croup, or laryngotracheobronchitis, which is characterized by a "barking" cough but can also include laryngitis.

Beyond viruses, laryngitis can be caused in rare cases by bacteria and fungi. Bacterial laryngitis can be a complication of viral laryngitis, occurring about 7 days into the illness. The most common bacteria involved are *S. pneumoniae, H. influenzae,* and *M. catarrhalis.* Fungal laryngitis is probably rarer but should be considered in patients who are immunosuppressed or who have recently been treated with antibacterial drugs.

Noninfectious causes of laryngitis include vocal trauma (e.g., due to yelling, screaming, or loud singing), inhalation injuries, allergies, gastroesophageal reflux disease (laryngopharyngeal reflux), asthma, and pollution. Immunosuppressed patients are at risk for infections with herpesvirus, HIV, and coxsackievirus. Smokers are at elevated risk for malignancy and other infections.

Laryngitis is characterized by a raspy, hoarse, or breathy voice, sometimes progressing to a complete loss of voice. Laryngitis can have associated dry cough and anterior throat pain; patients often feel a need to clear their throats. The physical examination in patients who may have laryngitis should focus on the head, neck, and lungs, but the diagnosis of laryngitis is generally based on history. If visualization of the vocal cords is necessary, indirect examination with a mirror or flexible laryngoscopy usually shows erythema and edema of the vocal cords and surrounding structures.

TREATMENT

Laryngitis

Laryngitis is generally self-limited, usually lasting 3–7 days, but may last up to 14 days. Vocal rest is crucial. Airway humidification and hydration

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Red flags for emergency evaluation and monitoring include shortness of breath, stridor, dysphagia, odynophagia, drooling, and posturing that could indicate epiglottitis. Referral to an otolaryngologist should be considered for patients who rely on their voice for work, such as singers and

Access Provided by:
SILVERCHAIR

examination with a mirror or flexible laryngoscopy usually shows erythema and edema of the vocal cords and surrounding tissues.

TREATMENT

Laryngitis

Laryngitis is generally self-limited, usually lasting 3–7 days, but may last up to 14 days. Vocal rest is crucial. Airway humidification and hydration should help. Patients likely to have laryngopharyngeal reflux should avoid gastroesophageal reflux–inducing foods and behaviors and should take antireflux medications. In randomized controlled trials, antibiotics were not effective in decreasing objective symptoms of laryngitis.

Red flags for emergency evaluation and monitoring include shortness of breath, stridor, dysphagia, odynophagia, drooling, and posturing that could indicate epiglottitis. Referral to an otolaryngologist should be considered for patients who rely on their voice for work, such as singers and teachers. A history of smoking or weight loss should raise suspicion of malignancy. Symptoms lasting >3 weeks should prompt referral to an otolaryngologist or speech specialist.

FURTHER READING

Centor RM, Linder JA: Web exclusive. Annals on call—*Fusobacterium* pharyngitis debate. Ann Intern Med 171:OC1, 2019.

Chua KP et al: Appropriateness of outpatient antibiotic prescribing among privately insured US patients: ICD-10-CM based cross sectional study. BMJ 364:k5092, 2019.

Lieberthal AS et al: Clinical practice guideline: The diagnosis and management of acute otitis media. Pediatrics 131:e964, 2013.

Rowe TA, Linder JA: Novel approaches to decrease inappropriate ambulatory antibiotic use. Expert Rev Anti Infect Ther 17:511, 2019.

Sanchez GV et al: Core elements of outpatient antibiotic stewardship. MMWR Recomm Rep 65:1, 2016.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 36: Oral Manifestations of Disease

Samuel C. Durso

INTRODUCTION

As primary care physicians and consultants, internists are often asked to evaluate patients with disease of the oral soft tissues, teeth, and pharynx. Knowledge of the oral milieu and its unique structures is necessary to guide preventive services and recognize oral manifestations of local or systemic disease **(Chap. A3)**. Furthermore, internists frequently collaborate with dentists in the care of patients who have a variety of medical conditions that affect oral health or who undergo dental procedures that increase their risk of medical complications.

DISEASES OF THE TEETH AND PERIODONTAL STRUCTURES

Tooth formation begins during the sixth week of embryonic life and continues through 17 years of age. Teeth start to develop in utero and continue to develop until after the tooth erupts. Normally, all 20 deciduous teeth have erupted by age 3 and have been shed by age 13. Permanent teeth, eventually totaling 32, begin to erupt by age 6 and have completely erupted by age 14, though third molars ("wisdom teeth") may erupt later.

The erupted tooth consists of the visible *crown* covered with enamel and the root submerged below the gum line and covered with bonelike *cementum*. *Dentin*, a material that is denser than bone and exquisitely sensitive to pain, forms the majority of the tooth substance, surrounding a core of myxomatous *pulp* containing the vascular and nerve supply. The tooth is held firmly in the alveolar socket by the *periodontium*, supporting structures that consist of the gingivae, alveolar bone, cementum, and periodontal ligament. The periodontal ligament tenaciously binds the tooth's cementum to the alveolar bone. Above this ligament is a collar of attached gingiva just below the crown. A few millimeters of unattached or free gingiva (1–3 mm) overlap the base of the crown, forming a shallow sulcus along the gum-tooth margin.

Dental Caries, Pulpal and Periapical Disease, and Complications

Dental caries usually begin asymptomatically as a destructive infectious process of the enamel. Bacteria—principally *Streptococcus mutans*—colonize the organic buffering biofilm (*plaque*) on the tooth surface. If not removed by brushing or by the natural cleansing and antibacterial action of saliva, bacterial acids can demineralize the enamel. Fissures and pits on the occlusal surfaces are the most frequent sites of early decay. Surfaces between the teeth, adjacent to tooth restorations and exposed roots, are also vulnerable, particularly as individuals age. Over time, dental caries extend to the underlying dentin, leading to cavitation of the enamel. Without management, the caries will penetrate to the tooth pulp, producing *acute pulpitis*. At this stage, when the pulp infection is limited, the tooth may become sensitive to percussion and to hot or cold, and pain resolves immediately when the irritating stimulus is removed. Should the infection spread throughout the pulp, *irreversible pulpitis* occurs, leading to *pulp necrosis*. At this later stage, pain can be severe and has a sharp or throbbing visceral quality that may be worse when the patient lies down. Once pulp necrosis is complete, pain may be constant or intermittent, but cold sensitivity is lost.

Treatment of caries involves removal of the softened and infected hard tissue and restoration of the tooth structure with silver amalgam, glass ionomer, composite resin, or gold. Once irreversible pulpitis occurs, root canal therapy becomes necessary; removal of the contents of the pulp chamber and root canal is followed by thorough cleaning and filling with an inert material. Alternatively, the tooth may be extracted.

Pulpal infection leads to *periapical abscess* formation, which can produce pain on chewing. If the infection is mild and chronic, a *periapical granuloma* or eventually a *periapical cyst* forms, either of which produces radiolucency at the root apex. When unchecked, a periapical abscess can erode into the alveolar bone, producing osteomyelitis; penetrate and drain through the gingivae, producing a parulis (gumboil); or track along deep fascial planes, producing virulent cellulitis (Ludwig's angina) involving the submandibular space and floor of the mouth **(Chap. 177)**. Elderly patients, patients with diabetes mellitus, and patients taking glucocorticoids may experience little or no pain or fever as these complications develop.

Periodontal Disease

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

appropriate daily oral hygiene, including periodic professional cleaning. Left undisturbed, chronic inflammation can ensue and produce hyperemia of the free and attached gingivae (*gingivitis*), which then typically bleed with brushing. If this issue is ignored, severe *periodontitis* can develop, leading to

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Periodontal disease and dental caries are the primary causes of tooth loss. Like dental caries, chronic infection of the gingiva and anchoring structures of the tooth begins with formation of bacterial plaque. The process begins at the gum line. Plaque and *calculus* (calcified plaque) are preventable by appropriate daily oral hygiene, including periodic professional cleaning. Left undisturbed, chronic inflammation can ensue and produce hyperemia of the free and attached gingivae (*gingivitis*), which then typically bleed with brushing. If this issue is ignored, severe *periodontitis* can develop, leading to deepening of the physiologic sulcus and destruction of the periodontal ligament. Gingival pockets develop around the teeth. As the periodontium (including the supporting bone) is destroyed, the teeth loosen. A role for chronic inflammation due to chronic periodontal disease in promoting coronary heart disease and stroke has been proposed. Epidemiologic studies have demonstrated a moderate but significant association between chronic periodontal inflammation and atherogenesis, though a causal role remains unproven.

Acute and aggressive forms of periodontal disease are less common than the chronic forms described above. However, if the host is stressed or exposed to a new pathogen, rapidly progressive and destructive disease of the periodontal tissue can occur. A virulent example is *acute necrotizing ulcerative gingivitis*. The presentation includes sudden gingival inflammation, ulceration, bleeding, interdental gingival necrosis, and fetid halitosis. *Localized juvenile periodontitis*, which is seen in adolescents, is particularly destructive and appears to be associated with impaired neutrophil chemotaxis. *AIDS-related periodontitis* resembles acute necrotizing ulcerative gingivitis in some patients and a more destructive form of adult chronic periodontitis in others. It may also produce a gangrene-like destructive process of the oral soft tissues and bone that resembles *noma*, an infectious condition seen in severely malnourished children in developing nations.

Prevention of Tooth Decay and Periodontal Infection

Despite the reduced prevalences of dental caries and periodontal disease in the United States (due in large part to water fluoridation and improved dental care, respectively), both diseases constitute a major public health problem worldwide, particularly in certain groups. The internist should promote preventive dental care and hygiene as part of health maintenance. Populations at high risk for dental caries and periodontal disease include those with hyposalivation and/or xerostomia, diabetics, alcoholics, tobacco users, persons with Down syndrome, and those with gingival hyperplasia. Furthermore, patients lacking access to dental care (e.g., as a result of low socioeconomic status) and patients with a reduced ability to provide self-care (e.g., individuals with disabilities, nursing home residents, and persons with dementia or upper-extremity disability) suffer at a disproportionate rate. It is important to provide counseling regarding regular dental hygiene and professional cleaning, use of fluoride-containing toothpaste, professional fluoride treatments, and (for patients with limited dexterity) use of electric toothbrushes and also to instruct persons caring for those who are not capable of self-care. Cost, fear of dental care, and differences in language and culture create barriers that prevent some people from seeking preventive dental services.

Developmental and Systemic Disease Affecting the Teeth and Periodontium

In addition to posing cosmetic issues, *malocclusion*, the most common developmental oral problem, can interfere with mastication unless corrected through orthodontic and surgical techniques. Impacted third molars are common and can become infected or erupt into an insufficient space. Acquired prognathism due to *acromegaly* may also lead to malocclusion, as may deformity of the maxilla and mandible due to *Paget's disease* of the bone. Delayed tooth eruption, a receding chin, and a protruding tongue are occasional features of *cretinism* and *hypopituitarism*. Congenital syphilis produces tapering, notched (*Hutchinson's*) incisors and finely nodular (*mulberry*) molar crowns. *Enamel hypoplasia* results in crown defects ranging from pits to deep fissures of primary or permanent teeth. Intrauterine infection (syphilis, rubella), vitamin deficiency (A, C, or D), disorders of calcium metabolism (malabsorption, vitamin D–resistant rickets, hypoparathyroidism), prematurity, high fever, and rare inherited defects (*amelogenesis imperfecta*) are all causes. Tetracycline, given in sufficiently high doses during the first 8 years of life, may produce enamel hypoplasia and discoloration. Doxycycline does not cause permanent tooth staining in children despite warnings included for all tetracycline-class antibiotics. Exposure to endogenous pigments can discolor developing teeth; etiologies include *erythroblastosis fetalis* (green or bluish-black), congenital liver disease (green or yellow-brown), and porphyria (red or brown that fluoresces with ultraviolet light). *Mottled enamel* occurs if excessive fluoride is ingested during development. Worn enamel is seen with age, bruxism, or excessive acid exposure (e.g., chronic gastric reflux or bulimia). Celiac disease is associated with nonspecific enamel defects in children but not in adults.

Total or partial tooth loss resulting from periodontitis is seen with cyclic neutropenia, Papillon-Lefévre syndrome, Chédiak-Higashi syndrome, and leukemia. Rapid focal tooth loosening is most often due to infection, but rarer causes include Langerhans cell histiocytosis, Ewing's sarcoma, osteosarcoma, and Burkitt's lymphoma. Early loss of primary teeth is a feature of *hypophosphatasia*, a rare congenital error of metabolism.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Discontinuation of the medication may reverse the drug-induced form, although surgery may be needed to control both of the latter entities. *Linear*

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

osteosarcoma, and Burkitt's lymphoma. Early loss of primary teeth is a feature of *hypophosphatasia*, a rare congenital error of metabolism.

Pregnancy may produce gingivitis and localized *pyogenic granulomas*. Severe periodontal disease occurs in uncontrolled diabetes mellitus. *Drug-induced gingival overgrowth* may be caused by anticonvulsants, calcium channel blockers, and immunosuppressants, although excellent daily oral care can prevent or reduce its occurrence. *Idiopathic familial gingival fibromatosis* and several syndrome-related disorders cause similar conditions. Discontinuation of the medication may reverse the drug-induced form, although surgery may be needed to control both of the latter entities. *Linear gingival erythema* is variably seen in patients with advanced HIV infection and probably represents immune deficiency and decreased neutrophil activity. Diffuse or focal gingival swelling may be a feature of early or late acute myelomonocytic leukemia as well as of other lymphoproliferative disorders. A rare but pathognomonic sign of granulomatosis with polyangiitis is a red-purplish, granular gingivitis (*strawberry gums*).

DISEASES OF THE ORAL MUCOSA

Infections

Most oral mucosal diseases involve microorganisms **(Table 36-1)**.

TABLE 36-1

Vesicular, Bullous, or Ulcerative Lesions of the Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Viral Diseases			
Primary acute herpetic gingivostomatitis (HSV type 1; rarely type 2)	Lip and oral mucosa (buccal, gingival, lingual mucosa)	Labial vesicles that rupture and crust, and intraoral vesicles that quickly ulcerate; extremely painful; acute gingivitis, fever, malaise, foul odor, and cervical lymphadenopathy; occurs primarily in infants, children, and young adults	Heals spontaneously in 10–14 days; unless secondarily infected, lesions lasting >3 weeks are not due to primary HSV infection
Recurrent herpes labialis	Mucocutaneous junction of lip, perioral skin	Eruption of groups of vesicles that may coalesce, then rupture and crust; painful to pressure or spicy foods	Lasts ~1 week, but condition may be prolonged if secondarily infected; if severe, topical or oral antiviral treatment may reduce healing time
Recurrent intraoral herpes simplex	Palate and gingiva	Small vesicles on keratinized epithelium that rupture and coalesce; painful	Heals spontaneously in ~1 week; if severe, topical or oral antiviral treatment may reduce healing time
Chickenpox (VZV)	Gingiva and oral mucosa	Skin lesions may be accompanied by small vesicles on oral mucosa that rupture to form shallow ulcers; may coalesce to form large bullous lesions that ulcerate; mucosa may have generalized erythema	Lesions heal spontaneously within 2 weeks
Herpes zoster (VZV reactivation)	Cheek, tongue, gingiva, or palate	Unilateral vesicular eruptions and ulceration in linear pattern following sensory distribution of trigeminal nerve or one of its branches	Gradual healing without scarring unless secondarily infected; postherpetic neuralgia is common; oral acyclovir, famciclovir, or valacyclovir reduces healing time and postherpetic neuralgia
Infectious mononucleosis (Epstein-Barr	Oral mucosa	Fatigue, sore throat, malaise, fever, and cervical lymphadenopathy; numerous small ulcers usually appear several days before lymphadenopathy;	Oral lesions disappear during convalescence; no treatment is given, though glucocorticoids are indicated if tonsillar swelling compromises
Herpangina (coxsackievirus A;	Oral mucosa, pharynx, tongue	Sudden onset of fever, sore throat, and oropharyngeal vesicles, usually in children <4 years	Incubation period of 2–9 days; fever for 1–4 days; recovery uneventful

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

(Epstein-Barr virus)		appear several days before lymphadenopathy; gingival bleeding and multiple petechiae at junction of hard and soft palates	are indicated if tonsillar swelling compromises the airway
Herpangina (coxsackievirus A; also possibly coxsackievirus B and echovirus)	Oral mucosa, pharynx, tongue	Sudden onset of fever, sore throat, and oropharyngeal vesicles, usually in children <4 years old, during summer months; diffuse pharyngeal congestion and vesicles (1–2 mm), grayish-white surrounded by red areola; vesicles enlarge and ulcerate	Incubation period of 2–9 days; fever for 1–4 days; recovery uneventful
Hand-foot-and-mouth disease (most commonly coxsackievirus A16)	Oral mucosa, pharynx, palms, and soles	Fever, malaise, headache with oropharyngeal vesicles that become painful, shallow ulcers; highly infectious; usually affects children under age 10	Incubation period 2–18 days; lesions heal spontaneously in 2–4 weeks
Primary HIV infection	Gingiva, palate, and pharynx	Acute gingivitis and oropharyngeal ulceration, associated with febrile illness resembling mononucleosis and including lymphadenopathy	Followed by HIV seroconversion, asymptomatic HIV infection, and usually ultimately by HIV disease
Bacterial or Fungal Diseases			
Acute necrotizing ulcerative gingivitis ("trench mouth")	Gingiva	Painful, bleeding gingiva characterized by necrosis and ulceration of gingival papillae and margins plus lymphadenopathy and foul breath	Debridement and diluted (1:3) peroxide lavage provide relief within 24 h; antibiotics in acutely ill patients; relapse may occur
Prenatal (congenital) syphilis	Palate, jaws, tongue, and teeth	Gummatous involvement of palate, jaws, and facial bones; Hutchinson's incisors, mulberry molars, glossitis, mucous patches, and fissures at corner of mouth	Tooth deformities in permanent dentition irreversible
Primary syphilis (chancre)	Lesion appearing where organism enters body; may occur on lips, tongue, or tonsillar area	Small papule developing rapidly into a large, painless ulcer with indurated border; unilateral lymphadenopathy; chancre and lymph nodes containing spirochetes; serologic tests positive by third to fourth weeks	Healing of chancre in 1–2 months, followed by secondary syphilis in 6–8 weeks
Secondary syphilis	Oral mucosa frequently involved with mucous patches, which occur primarily on palate and also at commissures of mouth	Maculopapular lesions of oral mucosa, 5–10 mm in diameter with central ulceration covered by grayish membrane; eruptions occurring on various mucosal surfaces and skin, accompanied by fever, malaise, and sore throat	Lesions may persist from several weeks to a year
Tertiary syphilis	Palate and tongue	Gummatous infiltration of palate or tongue followed by ulceration and fibrosis; atrophy of tongue papillae produces characteristic bald tongue and glossitis	Gumma may destroy palate, causing complete perforation
Gonorrhea	Lesions may occur in mouth at site of	Most pharyngeal infection is asymptomatic; may produce burning or itching sensation; oropharynx	More difficult to eradicate than urogenital infection, though pharyngitis usually resolves

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Tuberculosis	Tongue, tonsillar area,	Painless, solitary, 1- to 5-cm, irregular ulcer covered	Autoinoculation from pulmonary infection is

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Tertiary syphilis	Palate and tongue	Gummatous infiltration of palate or tongue followed	Gumma may destroy palate, causing
Gonorrhea	Lesions may occur in mouth at site of inoculation or secondarily by hematogenous spread from a primary focus	Most pharyngeal infection is asymptomatic; may produce burning or itching sensation; oropharynx and tonsils may be ulcerated and erythematous; saliva viscous and fetid	More difficult to eradicate than urogenital infection, though pharyngitis usually resolves with appropriate antimicrobial treatment
Tuberculosis	Tongue, tonsillar area, soft palate	Painless, solitary, 1- to 5-cm, irregular ulcer covered with persistent exudate; ulcer has firm undermined border	Autoinoculation from pulmonary infection is usual; lesions resolve with appropriate antimicrobial therapy
Cervicofacial actinomycosis	Swellings in region of face, neck, and floor of mouth	Infection may be associated with extraction, jaw fracture, or eruption of molar tooth; in acute form, resembles acute pyogenic abscess, but contains yellow “sulfur granules” (gram-positive mycelia and their hyphae)	Typically, swelling is hard and grows painlessly; multiple abscesses with draining tracts develop; penicillin first choice; surgery usually necessary
Histoplasmosis	Any area of the mouth, particularly tongue, gingiva, or palate	Nodular, verrucous, or granulomatous lesions; ulcers are indurated and painful; usual source hematogenous or pulmonary, but may be primary	Systemic antifungal therapy necessary
Candidiasis[a]			
Dermatologic Diseases			
Mucous membrane pemphigoid	Typically produces marked gingival erythema and ulceration; other areas of oral cavity, esophagus, and vagina may be affected	Painful, grayish-white collapsed vesicles or bullae of full-thickness epithelium with peripheral erythematous zone; gingival lesions desquamate, leaving ulcerated area	Protracted course with remissions and exacerbations; involvement of different sites develops slowly; glucocorticoids may temporarily reduce symptoms but do not control disease
EM minor and EM major (Stevens-Johnson syndrome)	Primarily oral mucosa and skin of hands and feet	Intraoral ruptured bullae surrounded by inflammatory area; lips may show hemorrhagic crusts; “iris” or “target” lesion on skin is pathognomonic; patient may have severe signs of toxicity	Onset very rapid; usually idiopathic, but may be associated with trigger such as drug reaction; condition may last 3–6 weeks; mortality rate for untreated EM major is 5–15%
Pemphigus vulgaris	Oral mucosa and skin; sites of mechanical trauma (soft/hard palate, frenulum, lips, buccal mucosa)	Usually (>70%) presents with oral lesions; fragile, ruptured bullae and ulcerated oral areas; mostly in older adults	With repeated occurrence of bullae, toxicity may lead to cachexia, infection, and death within 2 years; often controllable with oral glucocorticoids
Lichen planus	Oral mucosa and skin	White striae in mouth; purplish nodules on skin at sites of friction; occasionally causes oral mucosal ulcers and erosive gingivitis	White striae alone usually asymptomatic; erosive lesions often difficult to treat, but may respond to glucocorticoids
Other Conditions			
Recurrent aphthous ulcers	Usually on nonkeratinized oral mucosa (buccal and labial mucosa, floor of	Single or clustered painful ulcers with surrounding erythematous border; lesions may be 1–2 mm in diameter in crops (herpetiform), 1–5 mm (minor), or	Lesions heal in 1–2 weeks but may recur monthly or several times a year; protective barrier with benzocaine and topical

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Behçet’s	Oral mucosa, eyes,	Multiple aphthous ulcers in mouth; inflammatory	Oral lesions often first manifestation; persist

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

aphthous ulcers	oral mucosa (buccal and labial mucosa, floor of mouth, soft palate, lateral and ventral tongue)	erythematous border; lesions may be 1–2 mm in diameter in crops (herpetiform), 1–5 mm (minor), or 5–15 mm (major)	monthly or several times a year; protective barrier with benzocaine and topical glucocorticoids relieve symptoms; systemic glucocorticoids may be needed in severe cases
Behçet's syndrome	Oral mucosa, eyes, genitalia, gut, and CNS	Multiple aphthous ulcers in mouth; inflammatory ocular changes, ulcerative lesions on genitalia; inflammatory bowel disease and CNS disease	Oral lesions often first manifestation; persist several weeks and heal without scarring
Traumatic ulcers	Anywhere on oral mucosa; dentures frequently responsible for ulcers in vestibule	Localized, discrete ulcerated lesions with red border; produced by accidental biting of mucosa, penetration by foreign object, or chronic irritation by dentures	Lesions usually heal in 7–10 days when irritant is removed, unless secondarily infected
Squamous cell carcinoma	Any area of mouth, most commonly on lower lip, lateral borders of tongue, and floor of mouth	Red, white, or red and white ulcer with elevated or indurated border; failure to heal; pain not prominent in early lesions	Invades and destroys underlying tissues; frequently metastasizes to regional lymph nodes
Acute myeloid leukemia (usually monocytic)	Gingiva	Gingival swelling and superficial ulceration followed by hyperplasia of gingiva with extensive necrosis and hemorrhage; deep ulcers may occur elsewhere on mucosa, complicated by secondary infection	Usually responds to systemic treatment of leukemia; occasionally requires local irradiation
Lymphoma	Gingiva, tongue, palate, and tonsillar area	Elevated, ulcerated area that may proliferate rapidly, giving appearance of traumatic inflammation	Fatal if untreated; may indicate underlying HIV infection
Chemical or thermal burns	Any area in mouth	White slough due to contact with corrosive agents (e.g., aspirin, hot cheese) applied locally; removal of slough leaves raw, painful surface	Lesion heals in several weeks if not secondarily infected

[a]See Table 36-3.

Abbreviations: CNS, central nervous system; EM, erythema multiforme; HSV, herpes simplex virus; VZV, varicella-zoster virus.

Pigmented Lesions

See **Table 36-2**.

TABLE 36-2

Pigmented Lesions of the Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Oral melanotic macule	Any area of mouth	Discrete or diffuse, localized, brown to black macule	Remains indefinitely; no growth
Diffuse melanin pigmentation	Any area of mouth	Diffuse pale to dark-brown pigmentation; may be physiologic ("racial") or due to smoking	Remains indefinitely
Malignant melanoma	Any area of	Can be flat and diffuse, painless, brown to black; or can	Expands and invades early; metastasis

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Diffuse melanin pigmentation	Any area of mouth	Diffuse pale to dark-brown pigmentation; may be physiologic ("racial") or due to smoking	Remains indefinitely
Nevi	Any area of mouth	Discrete, localized, brown to black pigmentation	Remains indefinitely
Malignant melanoma	Any area of mouth	Can be flat and diffuse, painless, brown to black; or can be raised and nodular	Expands and invades early; metastasis leads to death
Addison's disease	Any area of mouth, but mostly buccal mucosa	Blotches or spots of bluish-black to dark-brown pigmentation occurring early in disease, accompanied by diffuse pigmentation of skin; other symptoms of adrenal insufficiency	Condition controlled by adrenal steroid replacement
Peutz-Jeghers syndrome	Any area of mouth	Dark-brown spots on lips, buccal mucosa, with characteristic distribution of pigment around lips, nose, and eyes and on hands; concomitant intestinal polyposis	Oral pigmented lesions remain indefinitely; gastrointestinal polyps may become malignant
Drug ingestion (neuroleptics, oral contraceptives, minocycline, zidovudine, quinine derivatives)	Any area of mouth	Brown, black, or gray areas of pigmentation	Gradually disappears following cessation of drug intake
Amalgam tattoo	Gingiva and alveolar mucosa	Small blue-black pigmented areas associated with embedded amalgam particles in soft tissues; may show up on radiographs as radiopaque particles in some cases	Remains indefinitely
Heavy metal pigmentation (bismuth, mercury, lead)	Gingival margin	Thin blue-black pigmented line along gingival margin; rarely seen except in children exposed to lead-based paint	Indicative of systemic absorption; no significance for oral health
Black hairy tongue	Dorsum of tongue	Elongation of filiform papillae of tongue, which become stained by coffee, tea, tobacco, or pigmented bacteria	Improves within 1–2 weeks with gentle brushing of tongue or (if due to bacterial overgrowth) discontinuation of antibiotic
Fordyce spots	Buccal and labial mucosa	Numerous small yellowish spots just beneath mucosal surface; no symptoms; due to hyperplasia of sebaceous glands	Benign; remains without apparent change
Kaposi's sarcoma	Palate most common, but may occur at any other site	Red or blue plaques of variable size and shape; often enlarge, become nodular, and may ulcerate	Usually indicative of HIV infection or non-Hodgkin's lymphoma; rarely fatal, but may require treatment for comfort or cosmesis
Mucous retention cysts	Buccal and labial mucosa	Bluish, clear fluid–filled cyst due to extravasated mucus from injured minor salivary gland	Benign; painless unless traumatized; may be removed surgically

Dermatologic Diseases

See Tables 36-1, 36-2, and **36-3** and **Chaps. 56–61**.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

CONDITION	USUAL	CLINICAL FEATURES	COURSE

See Tables 36-1, 36-2, and 36-3 and Chaps. 36–61.

TABLE 36-3

White Lesions of Oral Mucosa

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Lichen planus	Buccal mucosa, tongue, gingiva, and lips; skin	Striae, white plaques, red areas, ulcers in mouth; purplish papules on skin; may be asymptomatic, sore, or painful; lichenoid drug reactions may look similar	Protracted; responds to topical glucocorticoids
White sponge nevus	Oral mucosa, vagina, anal mucosa	Painless white thickening of epithelium; adolescence/early adulthood onset; familial	Benign and permanent
Smoker's leukoplakia and smokeless tobacco lesions	Any area of oral mucosa, sometimes related to location of habit	White patch that may become firm, rough, or red-fissured and ulcerated; may become sore and painful but is usually painless	May or may not resolve with cessation of habit; 2% of patients develop squamous cell carcinoma; early biopsy essential
Erythroplakia with or without white patches	Floor of mouth commonly affected in men; tongue and buccal mucosa in women	Velvety, reddish plaque; occasionally mixed with white patches or smooth red areas	High risk of squamous cell cancer; early biopsy essential
Candidiasis	Any area in mouth	*Pseudomembranous type* ("thrush"): creamy white curdlike patches that reveal a raw, bleeding surface when scraped; found in sick infants, debilitated elderly patients receiving high-dose glucocorticoids or broad-spectrum antibiotics, and patients with AIDS	Responds favorably to antifungal therapy and correction of predisposing causes where possible
		Erythematous type: flat, red, sometimes sore areas in same groups of patients	Course same as for pseudomembranous type
		Candidal leukoplakia: nonremovable white thickening of epithelium due to *Candida*	Responds to prolonged antifungal therapy
		Angular cheilitis: sore fissures at corner of mouth	Responds to topical antifungal therapy
Hairy leukoplakia	Usually on lateral tongue, rarely elsewhere on oral mucosa	White areas ranging from small and flat to extensive accentuation of vertical folds; found in HIV carriers (all risk groups for AIDS)	Due to Epstein-Barr virus; responds to high-dose acyclovir but recurs; rarely causes discomfort unless secondarily infected with *Candida*
Warts (human papillomavirus)	Anywhere on skin and oral mucosa	Single or multiple papillary lesions with thick, white, keratinized surfaces containing many pointed projections; cauliflower lesions covered with normal-colored mucosa or multiple pink or pale bumps (focal epithelial hyperplasia)	Lesions grow rapidly and spread; squamous cell carcinoma must be ruled out with biopsy; excision or laser therapy; may regress in HIV-infected patients receiving

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

CONDITION	USUAL LOCATION	CLINICAL FEATURES	COURSE
Lichen planus	Buccal mucosa, tongue, gingiva, and lips; skin	Striae, white plaques, red areas, ulcers in mouth; purplish papules on skin; may be asymptomatic, sore, or painful; lichenoid drug reactions may look similar	Protracted; responds to topical glucocorticoids
White sponge nevus	Oral mucosa, vagina, anal mucosa	Painless white thickening of epithelium; adolescence/early adulthood onset; familial	Benign and permanent
Smoker's leukoplakia and smokeless tobacco lesions	Any area of oral mucosa, sometimes related to location of habit	White patch that may become firm, rough, or red-fissured and ulcerated; may become sore and painful but is usually painless	May or may not resolve with cessation of habit; 2% of patients develop squamous cell carcinoma; early biopsy essential
Erythroplakia with or without white patches	Floor of mouth commonly affected in men; tongue and buccal mucosa in women	Velvety, reddish plaque; occasionally mixed with white patches or smooth red areas	High risk of squamous cell cancer; early biopsy essential
Candidiasis	Any area in mouth	*Pseudomembranous type* ("thrush"): creamy white curdlike patches that reveal a raw, bleeding surface when scraped; found in sick infants, debilitated elderly patients receiving high-dose glucocorticoids or broad-spectrum antibiotics, and patients with AIDS	Responds favorably to antifungal therapy and correction of predisposing causes where possible
		Erythematous type: flat, red, sometimes sore areas in same groups of patients	Course same as for pseudomembranous type
		Candidal leukoplakia: nonremovable white thickening of epithelium due to *Candida*	Responds to prolonged antifungal therapy
		Angular cheilitis: sore fissures at corner of mouth	Responds to topical antifungal therapy
Hairy leukoplakia	Usually on lateral tongue, rarely elsewhere on oral mucosa	White areas ranging from small and flat to extensive accentuation of vertical folds; found in HIV carriers (all risk groups for AIDS)	Due to Epstein-Barr virus; responds to high-dose acyclovir but recurs; rarely causes discomfort unless secondarily infected with *Candida*
Warts (human papillomavirus)	Anywhere on skin and oral mucosa	Single or multiple papillary lesions with thick, white, keratinized surfaces containing many pointed projections; cauliflower lesions covered with normal-colored mucosa or multiple pink or pale bumps (focal epithelial hyperplasia)	Lesions grow rapidly and spread; squamous cell carcinoma must be ruled out with biopsy; excision or laser therapy; may regress in HIV-infected patients receiving antiretroviral therapy

See Table 36-4.

TABLE 36-4

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Diseases of the Tongue

See **Table 36-4**.

TABLE 36-4

Alterations of the Tongue

TYPE OF CHANGE	CLINICAL FEATURES
Size or Morphology	
Macroglossia	Enlarged tongue that may be part of a syndrome found in developmental conditions such as Down syndrome, Simpson-Golabi-Behmel syndrome, or Beckwith-Wiedemann syndrome; may be due to tumor (hemangioma or lymphangioma), metabolic disease (e.g., primary amyloidosis), or endocrine disturbance (e.g., acromegaly or cretinism); may occur when all teeth are removed
Fissured ("scrotal") tongue	Dorsal surface and sides of tongue covered by painless shallow or deep fissures that may collect debris and become irritated
Median rhomboid glossitis	Congenital abnormality with ovoid, denuded area in median posterior portion of tongue; may be associated with candidiasis and may respond to antifungal treatment
Color	
"Geographic" tongue (benign migratory glossitis)	Asymptomatic inflammatory condition of tongue, with rapid loss and regrowth of filiform papillae leading to appearance of denuded red patches "wandering" across surface
Hairy tongue	Elongation of filiform papillae of medial dorsal surface area due to failure of keratin layer of papillae to desquamate normally; brownish-black coloration may be due to staining by tobacco, food, or chromogenic organisms
"Strawberry" and "raspberry" tongue	Appearance of tongue during scarlet fever due to hypertrophy of fungiform papillae as well as changes in filiform papillae
"Bald" tongue	Atrophy may be associated with xerostomia, pernicious anemia, iron-deficiency anemia, pellagra, or syphilis; may be accompanied by painful burning sensation; may be an expression of erythematous candidiasis and respond to antifungal treatment

HIV Disease and AIDS

See Tables 36-1, 36-2, 36-3, and **36-5**; **Chap. 202**.

TABLE 36-5

Oral Lesions Associated with HIV Infection

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Papules, nodules, plaques	Candidiasis (hyperplastic and pseudomembranous)[a] Condyloma acuminatum (human papillomavirus infection)

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

See Tables 36-1, 36-2, 36-3, and 36-5, Chap. 202.

TABLE 36-5

Oral Lesions Associated with HIV Infection

LESION MORPHOLOGY	ETIOLOGIES
Papules, nodules, plaques	Candidiasis (hyperplastic and pseudomembranous)[a] Condyloma acuminatum (human papillomavirus infection) Squamous cell carcinoma (preinvasive and invasive) Non-Hodgkin's lymphoma[a] Hairy leukoplakia[a]
Ulcers	Recurrent aphthous ulcers[a] Angular cheilitis Squamous cell carcinoma Acute necrotizing ulcerative gingivitis[a] Necrotizing ulcerative periodontitis[a] Necrotizing ulcerative stomatitis Non-Hodgkin's lymphoma[a] Viral infection (herpes simplex, herpes zoster, cytomegalovirus infection) Infection caused by *Mycobacterium tuberculosis* or *Mycobacterium avium-intracellulare* Fungal infection (histoplasmosis, cryptococcosis, candidiasis, geotrichosis, aspergillosis) Bacterial infection (*Escherichia coli, Enterobacter cloacae, Klebsiella pneumoniae, Pseudomonas aeruginosa*) Drug reactions (single or multiple ulcers)
Pigmented lesions	Kaposi's sarcoma[a] Bacillary angiomatosis (skin and visceral lesions more common than oral) Zidovudine pigmentation (skin, nails, and occasionally oral mucosa) Addison's disease
Miscellaneous	Linear gingival erythema[a]

[a]Strongly associated with HIV infection.

Ulcers

Ulceration is the most common oral mucosal lesion. Although there are many causes, the host and the pattern of lesions, including the presence of organ system features, narrow the differential diagnosis (Table 36-1). Most acute ulcers are painful and self-limited. Recurrent aphthous ulcers and herpes simplex account for the majority. Persistent and deep aphthous ulcers can be idiopathic or can accompany HIV/AIDS. Aphthous lesions are often the presenting symptom in *Behçet's syndrome* (Chap. 364). Similar-appearing, though less painful, lesions may occur in reactive arthritis, and aphthous ulcers are occasionally present during phases of *discoid* or *systemic lupus erythematosus* (Chap. 360). Aphthous-like ulcers are seen in *Crohn's disease* (Chap. 326), but, unlike the common aphthous variety, they may exhibit granulomatous inflammation on histologic examination. Recurrent aphthae are more prevalent in patients with *celiac disease* and have been reported to remit with elimination of gluten.

Of major concern are chronic, relatively painless ulcers and mixed red/white patches (erythroplakia and leukoplakia) of >2 weeks' duration. Squamous cell carcinoma and premalignant dysplasia should be considered early and a diagnostic biopsy performed. This awareness and this procedure are critically important because early-stage malignancy is vastly more treatable than late-stage disease. High-risk sites include the lower lip, floor of the mouth, ventral and lateral tongue, and soft palate–tonsillar pillar complex. Significant risk factors for oral cancer in Western countries include sun

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

clinical features and performing a biopsy of the lesion. The syphilitic chancre is typically painless and therefore easily missed. Regional lymphadenopathy is invariably present. The syphilitic etiology is confirmed with appropriate bacterial and serologic tests.

mouth, ventral and lateral tongue, and soft palate–tonsillar pillar complex. Significant risk factors for oral cancer in Western countries include sun exposure (lower lip), tobacco and alcohol use, and human papillomavirus infection. In India and some other Asian countries, smokeless tobacco mixed with betel nut, slaked lime, and spices is a common cause of oral cancer. Rarer causes of chronic oral ulcer, such as tuberculosis, fungal infection, granulomatosis with polyangiitis, and midline granuloma, may look identical to carcinoma. Making the correct diagnosis depends on recognizing other clinical features and performing a biopsy of the lesion. The syphilitic chancre is typically painless and therefore easily missed. Regional lymphadenopathy is invariably present. The syphilitic etiology is confirmed with appropriate bacterial and serologic tests.

Disorders of mucosal fragility often produce painful oral ulcers that fail to heal within 2 weeks. *Mucous membrane pemphigoid* and *pemphigus vulgaris* are the major acquired disorders. While their clinical features are often distinctive, a biopsy or immunohistochemical examination should be performed to diagnose these entities and to distinguish them from *lichen planus* and drug reactions.

Hematologic and Nutritional Disease

Internists are more likely to encounter patients with acquired, rather than congenital, bleeding disorders. Bleeding should stop 15 min after minor trauma and within an hour after tooth extraction if local pressure is applied. More prolonged bleeding, if not due to continued injury or rupture of a large vessel, should lead to investigation for a clotting abnormality. In addition to bleeding, petechiae and ecchymoses are prone to occur at the vibrating line between the soft and hard palates in patients with platelet dysfunction or thrombocytopenia.

All forms of leukemia, but particularly *acute myelomonocytic leukemia*, can produce gingival bleeding, ulcers, and gingival enlargement. Oral ulcers are a feature of agranulocytosis, and ulcers and mucositis are often severe complications of chemotherapy and radiation therapy for hematologic and other malignancies. *Plummer-Vinson syndrome* (iron deficiency, angular stomatitis, glossitis, and dysphagia) raises the risk of oral squamous cell cancer and esophageal cancer at the postcricoidal tissue web. Atrophic papillae and a red, burning tongue may occur with pernicious anemia. Deficiencies in B-group vitamins produce many of these same symptoms as well as oral ulceration and cheilosis. Consequences of *scurvy* include swollen, bleeding gums; ulcers; and loosening of the teeth.

NONDENTAL CAUSES OF ORAL PAIN

Most, but not all, oral pain emanates from inflamed or injured tooth pulp or periodontal tissues. Nonodontogenic causes are often overlooked. In most instances, toothache is predictable and proportional to the stimulus applied, and an identifiable condition (e.g., caries, abscess) is found. Local anesthesia eliminates pain originating from dental or periodontal structures, but not referred pains. The most common nondental source of pain is myofascial pain referred from muscles of mastication, which become tender and ache with increased use. Many sufferers exhibit *bruxism* (grinding of the teeth) secondary to stress and anxiety. *Temporomandibular joint disorder* is closely related. It affects both sexes, with a higher prevalence among women. Features include pain, limited mandibular movement, and temporomandibular joint sounds. The etiologies are complex; malocclusion does not play the primary role once attributed to it. *Osteoarthritis* is a common cause of masticatory pain. Anti-inflammatory medication, jaw rest, soft foods, and heat provide relief. The temporomandibular joint is involved in 50% of patients with *rheumatoid arthritis*, and its involvement is usually a late feature of severe disease. Bilateral preauricular pain, particularly in the morning, limits range of motion.

Migrainous neuralgia may be localized to the mouth. Episodes of pain and remission without an identifiable cause and a lack of relief with local anesthesia are important clues. *Trigeminal neuralgia* (*tic douloureux*) can involve the entire branch or part of the mandibular or maxillary branch of the fifth cranial nerve and can produce pain in one or a few teeth. Pain may occur spontaneously or may be triggered by touching the lip or gingiva, brushing the teeth, or chewing. *Glossopharyngeal neuralgia* produces similar acute neuropathic symptoms in the distribution of the ninth cranial nerve. Swallowing, sneezing, coughing, or pressure on the tragus of the ear triggers pain that is felt in the base of the tongue, pharynx, and soft palate and may be referred to the temporomandibular joint. *Neuritis* involving the maxillary and mandibular divisions of the trigeminal nerve (e.g., maxillary sinusitis, neuroma, and leukemic infiltrate) is distinguished from ordinary toothache by the neuropathic quality of the pain. Occasionally, *phantom pain* follows tooth extraction. Pain and hyperalgesia behind the ear and on the side of the face in the day or so before facial weakness develops often constitute the earliest symptom of *Bell's palsy*. Likewise, similar symptoms may precede visible lesions of herpes zoster infecting the seventh nerve (*Ramsey-Hunt syndrome*) or trigeminal nerve. *Postherpetic neuralgia* may follow either condition. *Coronary ischemia* may produce pain exclusively in the face and jaw; as in typical angina pectoris, this pain is usually reproducible with increased myocardial demand. Aching in several upper molar or premolar teeth that is unrelieved by anesthetizing the teeth may point to *maxillary sinusitis*.

Giant cell arteritis is notorious for producing headache, but it may also produce facial pain or sore throat without headache. Jaw and tongue claudication with chewing or talking is relatively common. Tongue infarction is rare. Patients with subacute thyroiditis often experience pain referred

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

low-grade *Candida* infection, food sensitivity, or subtle xerostomia) and predominantly affects postmenopausal women. The etiology may be neuropathic. Clonazepam, α-lipoic acid, and cognitive-behavioral therapy have benefited some patients. Some cases associated with an angiotensin-

claudication with chewing or talking is relatively common. Tongue infarction is rare. Patients with subacute thyroiditis often experience pain referred to the face or jaw before the tenderness of the thyroid gland and transient hyperthyroidism are appreciated.

"Burning mouth syndrome" (*glossodynia*) occurs in the absence of an identifiable cause (e.g., vitamin B_{12} deficiency, iron deficiency, diabetes mellitus, low-grade *Candida* infection, food sensitivity, or subtle xerostomia) and predominantly affects postmenopausal women. The etiology may be neuropathic. Clonazepam, α-lipoic acid, and cognitive-behavioral therapy have benefited some patients. Some cases associated with an angiotensin-converting enzyme inhibitor have remitted when treatment with the drug was discontinued.

DISEASES OF THE SALIVARY GLANDS

Saliva is essential to oral health. Its absence leads to dental caries, periodontal disease, and difficulties in wearing dental prostheses, masticating, and speaking. Its major components, water and mucin, serve as a cleansing solvent and lubricating fluid. In addition, saliva contains antimicrobial factors (e.g., lysozyme, lactoperoxidase, secretory IgA), epidermal growth factor, minerals, and buffering systems. The major salivary glands secrete intermittently in response to autonomic stimulation, which is high during a meal but low otherwise. Hundreds of minor glands in the lips and cheeks secrete mucus continuously throughout the day and night. Consequently, oral function becomes impaired when salivary function is reduced. The sensation of a dry mouth (*xerostomia*) is perceived when salivary flow is reduced by 50%. The most common etiology is medication, especially drugs with anticholinergic properties but also alpha and beta blockers, calcium channel blockers, and diuretics. Other causes include Sjögren's syndrome, chronic parotitis, salivary duct obstruction, diabetes mellitus, HIV/AIDS, and radiation therapy that includes the salivary glands in the field (e.g., for Hodgkin's lymphoma and for head and neck cancer). Management involves the elimination or limitation of drying medications, preventive dental care, and supplementation with oral liquid or salivary substitutes. Sugarless mints or chewing gum may stimulate salivary secretion if dysfunction is mild. When sufficient exocrine tissue remains, pilocarpine or cevimeline has been shown to increase secretions. Commercial saliva substitutes or gels relieve dryness. Fluoride supplementation is critical to prevent caries.

Sialolithiasis presents most often as painful swelling but in some instances as only swelling or only pain. Conservative therapy consists of local heat, massage, and hydration. Promotion of salivary secretion with mints or lemon drops may flush out small stones. Antibiotic treatment is necessary when bacterial infection in suspected. In adults, *acute bacterial parotitis* is typically unilateral and most commonly affects postoperative, dehydrated, and debilitated patients. *Staphylococcus aureus* (including methicillin-resistant strains) and anaerobic bacteria are the most common pathogens. Chronic bacterial *sialadenitis* results from lowered salivary secretion and recurrent bacterial infection. When suspected bacterial infection is not responsive to therapy, the differential diagnosis should be expanded to include benign and malignant neoplasms, lymphoproliferative disorders, Sjögren's syndrome, sarcoidosis, tuberculosis, lymphadenitis, actinomycosis, and granulomatosis with polyangiitis. Bilateral nontender parotid enlargement occurs with diabetes mellitus, cirrhosis, bulimia, HIV/AIDS, and drugs (e.g., iodide, propylthiouracil).

Pleomorphic adenoma composes two-thirds of all salivary neoplasms. The parotid is the principal salivary gland affected, and the tumor presents as a firm, slow-growing mass. Although this tumor is benign, its recurrence is common if resection is incomplete. Malignant tumors such as mucoepidermoid carcinoma, adenoid cystic carcinoma, and adenocarcinoma tend to grow relatively fast, depending upon grade. They may ulcerate and invade nerves, producing numbness and facial paralysis. Surgical resection is the primary treatment. Radiation therapy (particularly neutron-beam therapy) is used when surgery is not feasible and after resection for certain histologic types with a high risk of recurrence. Malignant salivary gland tumors have a 5-year survival rate of 94% when the stage is local and 35% when distant.

Dental Care for Medically Complex Patients

Routine dental care (e.g., uncomplicated extraction, scaling and cleaning, tooth restoration, and root canal) is remarkably safe. The most common concerns regarding care of dental patients with medical disease are excessive bleeding for patients taking anticoagulants, infection of the heart valves and prosthetic devices from hematogenous seeding by the oral flora, and cardiovascular complications resulting from vasopressors used with local anesthetics during dental treatment. Experience confirms that the risk of any of these complications is very low.

Patients undergoing tooth extraction or alveolar and gingival surgery rarely experience uncontrolled bleeding when warfarin anticoagulation is maintained within the therapeutic range currently recommended for prevention of venous thrombosis, atrial fibrillation, or mechanical heart valve. Embolic complications and death, however, have been reported during subtherapeutic anticoagulation. Therapeutic anticoagulation should be confirmed before and continued through the procedure. Likewise, low-dose aspirin (e.g., 81–325 mg) can safely be continued. For patients taking aspirin and another antiplatelet medication (e.g., clopidogrel), the decision to continue the second antiplatelet medication should be based on individual consideration of the risks of thrombosis and bleeding. The newer target-specific oral anticoagulants (dabigatran, apixaban, rivaroxaban, and edoxaban) are increasingly common use. Simple extractions of one to three teeth, periodontal surgery, abscess drainage, and implant

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Patients at risk for bacterial endocarditis **(Chap. 128)** should maintain optimal oral hygiene, including flossing, and have regular professional cleanings. Currently, guidelines recommend that prophylactic antibiotics be restricted to those patients at high risk for bacterial endocarditis who

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

individual consideration of the risks of thrombosis and bleeding. The newer target-specific oral anticoagulants (dabigatran, apixaban, rivaroxaban, and edoxaban) are in increasingly common use. Simple extractions of one to three teeth, periodontal surgery, abscess drainage, and implant positioning do not typically require interruption of therapy. More extensive surgery may necessitate delaying or holding a dose of the anticoagulant or more elaborate measures to manage the risk of thrombosis and bleeding.

Patients at risk for bacterial endocarditis **(Chap. 128)** should maintain optimal oral hygiene, including flossing, and have regular professional cleanings. Currently, guidelines recommend that prophylactic antibiotics be restricted to those patients at high risk for bacterial endocarditis who undergo dental and oral procedures involving significant manipulation of gingival or periapical tissue or penetration of the oral mucosa. If unexpected bleeding occurs, antibiotics given within 2 h after the procedure provide effective prophylaxis.

Hematogenous bacterial seeding from oral infection can undoubtedly produce late prosthetic-joint infection and therefore requires removal of the infected tissue (e.g., drainage, extraction, root canal) and appropriate antibiotic therapy. However, evidence that late prosthetic-joint infection follows routine dental procedures is lacking. For this reason, antibiotic prophylaxis is generally not recommended before oral surgery or oral mucosal manipulation for patients who have undergone joint replacement surgery. Exceptions to this may be considered for patients who have experienced joint replacement complications.

Concern often arises regarding the use of vasoconstrictors to treat patients with hypertension and heart disease. Vasoconstrictors enhance the depth and duration of local anesthesia, thus reducing the anesthetic dose and potential toxicity. If intravascular injection is avoided, 2% lidocaine with 1:100,000 epinephrine (limited to a total of 0.036 mg of epinephrine) can be used safely in patients with controlled hypertension and stable coronary heart disease, arrhythmia, or congestive heart failure. Precautions should be taken with patients taking tricyclic antidepressants and nonselective beta blockers because these drugs may potentiate the effect of epinephrine.

Elective dental treatments should be postponed for at least 1 month and preferably for 6 months after myocardial infarction, after which the risk of reinfarction is low provided the patient is medically stable (e.g., stable rhythm, stable angina, and no heart failure). Patients who have suffered a stroke should have elective dental care deferred for 9 months. In both situations, effective stress reduction requires good pain control, including the use of the minimal amount of vasoconstrictor necessary to provide good hemostasis and local anesthesia.

Bisphosphonate therapy is associated with *osteonecrosis* of the jaw. However, the risk with oral bisphosphonate therapy is very low. Most patients affected have received high-dose aminobisphosphonate therapy for multiple myeloma or metastatic breast cancer and have undergone tooth extraction or dental surgery. Intraoral lesions, of which two-thirds are painful, appear as exposed yellow-white hard bone involving the mandible or maxilla. Screening tests for determining risk of osteonecrosis are unreliable. Patients slated for aminobisphosphonate therapy should receive preventive dental care that reduces the risk of infection and the need for future dentoalveolar surgery.

Halitosis

Halitosis typically emanates from the oral cavity or nasal passages. Volatile sulfur compounds resulting from bacterial decay of food and cellular debris account for the malodor. Periodontal disease, caries, acute forms of gingivitis, poorly fitting dentures, oral abscess, and tongue coating are common causes. Treatment includes correcting poor hygiene, treating infection, and tongue brushing. Hyposalivation can produce and exacerbate halitosis. Pockets of decay in the tonsillar crypts, esophageal diverticulum, esophageal stasis (e.g., achalasia, stricture), sinusitis, and lung abscess account for some instances. A few systemic diseases produce distinctive odors: renal failure (ammoniacal), hepatic (fishy), and ketoacidosis (fruity). *Helicobacter pylori* gastritis can also produce ammoniacal breath. If a patient presents because of concern about halitosis but no odor is detectable, then pseudohalitosis or halitophobia must be considered.

Aging and Oral Health

While tooth loss and dental disease are not normal consequences of aging, a complex array of structural and functional changes that occur with age can affect oral health. Subtle changes in tooth structure (e.g., diminished pulp space and volume, sclerosis of dentinal tubules, and altered proportions of nerve and vascular pulp content) result in the elimination or diminution of pain sensitivity and a reduction in the reparative capacity of the teeth. In addition, age-associated fatty replacement of salivary acini may reduce physiologic reserve, thus increasing the risk of hyposalivation. In healthy older adults, there is minimal, if any, reduction in salivary flow.

Poor oral hygiene often results when general health fails or when patients lose manual dexterity and upper-extremity flexibility. This situation is particularly common among frail older adults and nursing home residents and must be emphasized because regular oral cleaning and dental care

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Periodontal disease, a leading cause of tooth loss, is indicated by loss of alveolar bone height. More than 90% of the U.S. population has some degree

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

particularly common among frail older adults and nursing home residents and must be emphasized because regular oral cleaning and dental care reduce the incidence of pneumonia and oral disease as well as the mortality risk in this population. Other risks for dental decay include limited lifetime fluoride exposure. Without assiduous care, decay can become quite advanced yet remain asymptomatic. Consequently, much of a tooth—or the entire tooth—can be destroyed before the patient is aware of the process.

Periodontal disease, a leading cause of tooth loss, is indicated by loss of alveolar bone height. More than 90% of the U.S. population has some degree of periodontal disease by age 50. Healthy adults who have not had significant alveolar bone loss by the sixth decade of life do not typically experience significant worsening with advancing age.

With the passing of those born in the first half of the twentieth century, complete edentulousness in the United States is becoming increasingly restricted to impoverished populations. When it is present, speech, mastication, and facial contours are dramatically affected. Edentulousness may also exacerbate obstructive sleep apnea, particularly in asymptomatic individuals who wear dentures. Dentures can improve verbal articulation and restore diminished facial contours. Mastication can also be restored; however, patients expecting dentures to facilitate oral intake are often disappointed. Accommodation to dentures requires a period of adjustment. Pain can result from friction or traumatic lesions produced by loose dentures. Poor fit and poor oral hygiene may permit the development of candidiasis. This fungal infection may be either asymptomatic or painful and is suggested by erythematous smooth or granular tissue conforming to an area covered by the appliance. Individuals with dentures and no natural teeth need regular (annual) professional oral examinations.

FURTHER READING

Durso SC: Interaction with other health team members in caring for elderly patients. Dent Clin North Am 49:377, 2005. [PubMed: 15755411]

Kaplovitch E, Dounaevskaia V: Treatment in the dental practice of the patient receiving anticoagulant therapy. J Am Dent Assoc 150:602, 2019. [PubMed: 31109636]

Weintraub JA et al: Improving nursing home residents' oral hygiene: Results of a cluster randomized intervention trial. J Am Med Dir Assoc 19:1086, 2018. [PubMed: 30471800]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 36: Oral Manifestations of Disease, Samuel C. Durso
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 38: Cough

Christopher H. Fanta

COUGH

Cough performs an essential protective function for human airways and lungs. Without an effective cough reflex, we are at risk for retained airway secretions and aspirated material predisposing to infection, atelectasis, and respiratory compromise. At the other extreme, excessive coughing can be exhausting; can be complicated by emesis, syncope, muscular pain, or rib fractures; can aggravate low back pain, abdominal or inguinal hernias, and urinary incontinence; and can be a major impediment to social interactions. Cough is often a clue to the presence of respiratory disease. In many instances, cough is an expected and accepted manifestation of disease, as in acute respiratory tract infection. However, persistent cough in the absence of other respiratory symptoms commonly causes patients to seek medical attention.

COUGH MECHANISM

Both chemical (e.g., capsaicin) and mechanical (e.g., mucus, particulates in air pollution) stimuli can initiate the cough reflex. Cationic channels (e.g., transient receptor potential channels) and adenosine triphosphate–activated ion channels (P2X3) function as sensory neuronal receptors, with signals transmitted centrally via Aδ (mechanosensory) and C fibers (chemosensory). Afferent nerve endings richly innervate the pharynx, larynx, and airways to the level of the terminal bronchioles and extend into the lung parenchyma. They are also located in the external auditory canal (the auricular branch of the vagus nerve, or Arnold's nerve) and in the esophagus. Sensory signals travel via the vagus and superior laryngeal nerves to a region of the brainstem in the nucleus tractus solitarius. Integrated neural networks process this input into a conscious sensation referred to as the "urge to cough." The efferent limb of the cough reflex involves a highly orchestrated series of involuntary muscular actions, with the potential for input from cortical pathways as well, making possible voluntary cough. The vocal cords adduct, leading to transient upper-airway occlusion. Expiratory muscles contract, generating positive intrathoracic pressures as high as 300 mmHg. With sudden release of the laryngeal contraction, rapid expiratory flows are generated, exceeding the normal "envelope" of maximal expiratory flow seen on the flow-volume curve **(Fig. 38-1)**. Bronchial smooth-muscle contraction together with dynamic compression of airways narrows airway lumens and maximizes the velocity of exhalation. The kinetic energy available to dislodge mucus from the inside of airway walls is directly proportional to the square of the velocity of expiratory airflow. A deep breath preceding a cough optimizes the function of the expiratory muscles; a series of repetitive coughs at successively lower lung volumes sweeps the point of maximal expiratory velocity progressively further into the lung periphery.

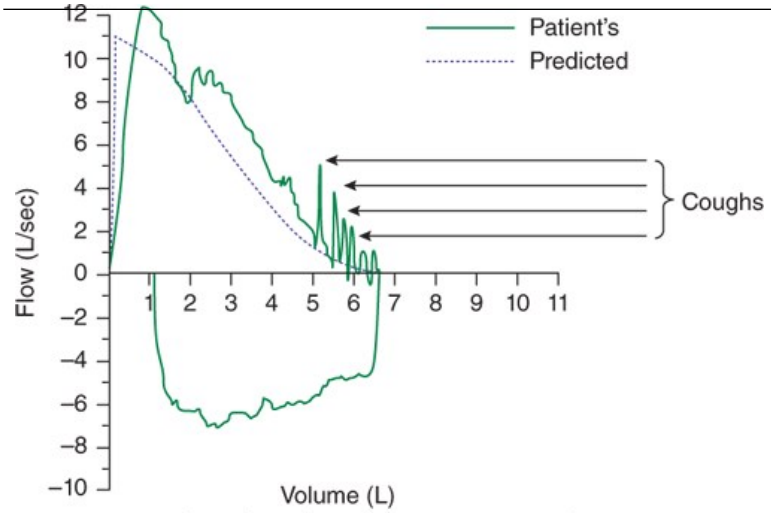
FIGURE 38-1

Flow-volume curve shows spikes of high expiratory flow achieved with cough.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 38: Cough, Christopher H. Fanta
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

of maximal expiratory velocity progressively farther into the lung periphery.

Flow-Volume Curve shows spikes of high expiratory flow achieved with cough.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

IMPAIRED COUGH

Weak or ineffective cough compromises the ability to clear lower respiratory tract secretions, predisposing to more serious infections and their sequelae. Weakness or paralysis of the expiratory (abdominal and intercostal) muscles and pain in the chest wall or abdomen are foremost on the list of causes of impaired cough **(Table 38-1)**. Cough strength is generally assessed qualitatively; peak expiratory flow or maximal expiratory pressure at the mouth can be used as a surrogate marker for cough strength. A variety of assistive devices and techniques have been developed to improve cough efficacy, running the gamut from simple (splinting of the abdominal muscles with a tightly held pillow to reduce postoperative pain while coughing) to complex (a mechanical cough-assist device supplied via face mask or tracheal tube that applies a cycle of positive pressure followed rapidly by negative pressure). Cough may fail to clear secretions completely despite a preserved ability to generate normal expiratory velocities; such failure may be due to abnormal airway secretions (e.g., abnormally viscous secretions of cystic fibrosis), ciliary dysfunction (e.g., primary ciliary dyskinesia), or structural abnormalities of the airways (e.g., tracheomalacia with excessive expiratory collapse of the trachea during cough).

TABLE 38-1

Causes of Impaired Cough and Airway Clearance

Respiratory muscle weakness
Chest wall or abdominal pain
Chest wall deformity (e.g., severe kyphoscoliosis)
Impaired glottic closure or tracheostomy
Central respiratory depression (e.g., anesthesia, sedation, or neurologic disease)
Abnormal airway secretions
Ciliary dysfunction
Tracheobronchomalacia
Bronchiectasis
Tracheal or bronchial stenoses

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 38: Cough, Christopher H. Fanta
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

SYMPTOMATIC COUGH

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

abnormalities of the airways (e.g., tracheomalacia with excessive expiratory collapse of the trachea during cough).

TABLE 38-1

Causes of Impaired Cough and Airway Clearance

Respiratory muscle weakness
Chest wall or abdominal pain
Chest wall deformity (e.g., severe kyphoscoliosis)
Impaired glottic closure or tracheostomy
Central respiratory depression (e.g., anesthesia, sedation, or neurologic disease)
Abnormal airway secretions
Ciliary dysfunction
Tracheobronchomalacia
Bronchiectasis
Tracheal or bronchial stenoses

SYMPTOMATIC COUGH

Cough may occur in the context of other respiratory symptoms that together point to a diagnosis; for example, cough accompanied by wheezing, shortness of breath, and chest tightness after exposure to a cat or other sources of allergens suggests asthma. At times, however, cough is the dominant or sole symptom of disease, and it may be of sufficient duration and severity that relief is sought. The duration of cough is a clue to its etiology, at least retrospectively. Acute cough (<3 weeks) is most commonly due to a respiratory tract infection, aspiration, or inhalation of noxious chemicals or smoke. Subacute cough (3–8 weeks in duration) is a common residuum of tracheobronchitis, as in pertussis or "postviral tussive syndrome." Chronic cough (>8 weeks) may be caused by a wide variety of cardiopulmonary diseases, including those of inflammatory, infectious, neoplastic, and cardiovascular etiologies. When initial assessment with chest examination and radiography is normal, cough-variant asthma, gastroesophageal reflux, rhinosinusitis with excessive nasopharyngeal drainage, and medications (angiotensin-converting enzyme [ACE] inhibitors) are the most common identifiable causes of chronic cough. In a long-time cigarette smoker, an early-morning, productive cough suggests chronic bronchitis. A dry, irritative cough that lingers for >2 months following one or more respiratory tract infections ("postbronchitic cough") is a very common cause of chronic cough, especially in the winter months. Chronic cough in the absence of identifiable etiology has been recognized with increasing frequency, is thought to be due to exaggerated neurologic signaling via sensory cough-reflex pathways, and is referred to as "chronic cough hypersensitivity syndrome."

ASSESSMENT OF CHRONIC COUGH

Except for our ability to detect the sound of excess airway secretions, details as to the resonance of the cough, its time of occurrence during the day, and the pattern of coughing (e.g., occurring in paroxysms) infrequently provide useful etiologic clues. Regardless of cause, cough often worsens upon first lying down at night, with talking, or with the hyperpnea of exercise; it frequently improves with sleep. An exception may involve the cough that occurs only with certain allergic exposures or exercise in cold air, as in asthma. Useful historical questions include what circumstances surrounded the onset of cough, what makes the cough better or worse, and whether the cough produces sputum.

The physical examination seeks clues suggesting the presence of cardiopulmonary disease, including findings such as wheezing or crackles on chest examination. Examination of the auditory canals and tympanic membranes (for irritation of the latter resulting in stimulation of Arnold's nerve), the

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 38: Cough, Christopher H. Fanta
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

symptoms and without detectable abnormalities on physical examination is long. It includes serious illnesses such as sarcoidosis or Hodgkin's disease in young adults, lung cancer in older patients, and (worldwide) pulmonary tuberculosis. An abnormal chest film prompts an evaluation aimed at

examination. Examination of the auditory canals and tympanic membranes (for irritation of the latter resulting in stimulation of Arnold's nerve), the nasal passageways (for rhinitis or polyps), and the nails (for clubbing) may also provide etiologic clues. Because cough can be a manifestation of a systemic disease such as sarcoidosis or vasculitis, a thorough general examination is likewise important.

In virtually all instances, evaluation of chronic cough merits a chest radiograph. The list of diseases that can cause persistent cough without other symptoms and without detectable abnormalities on physical examination is long. It includes serious illnesses such as sarcoidosis or Hodgkin's disease in young adults, lung cancer in older patients, and (worldwide) pulmonary tuberculosis. An abnormal chest film prompts an evaluation aimed at explaining the radiographic abnormality. In a patient with chronic productive cough, examination of expectorated sputum is warranted, because determining the cause of mucus hypersecretion is a crucial clue to etiology. Purulent-appearing sputum should be sent for routine bacterial culture and, in certain circumstances, mycobacterial culture as well. Cytologic examination of mucoid sputum may be useful to assess for malignancy and oropharyngeal aspiration and to distinguish neutrophilic from eosinophilic bronchitis. Expectoration of blood—whether streaks of blood, blood mixed with airway secretions, or pure blood—deserves a special approach to assessment and management (Chap. 39).

CHRONIC COUGH WITH A NORMAL CHEST RADIOGRAPH

It is commonly held that (alone or in combination) the use of an ACE inhibitor; postnasal drainage; gastroesophageal reflux; and asthma account for >90% of cases of chronic cough with a normal or noncontributory chest radiograph. However, clinical experience does not support this contention, and strict adherence to this concept discourages the search for alternative explanations by both clinicians and researchers. In recent years, the concept of a distinct "cough hypersensitivity syndrome" has emerged, emphasizing the putative role of sensitized sensory nerve endings and afferent neural pathways in causing chronic refractory cough, akin to chronic neuropathic pain. It presents with a dry or minimally productive cough and a tickle or sensitivity in the throat, made worse with talking, laughing, or exertion. It is more common in women than men and can last for years. Specific diagnostic criteria are lacking; the diagnosis is suspected when alternative etiologies are excluded by diagnostic testing or failed therapeutic trials. It is uncertain whether persistent daily coughing elicits an inflammatory response and is thereby self-perpetuating.

ACE inhibitor–induced cough occurs in 5–30% of patients taking these agents and is not dose-dependent. ACE metabolizes bradykinin and other tachykinins, such as substance P. The mechanism of ACE inhibitor–associated cough may involve sensitization of sensory nerve endings due to accumulation of bradykinin. Any patient with chronic unexplained cough who is taking an ACE inhibitor should have a trial period off the medication, regardless of the timing of the onset of cough relative to the initiation of ACE inhibitor therapy. In most instances, a safe alternative is available; angiotensin receptor blockers do not cause cough. Failure to observe a decrease in cough after 1 month off medication argues strongly against this etiology.

Postnasal drainage of any etiology can cause cough as a response to stimulation of sensory receptors of the cough-reflex pathway in the hypopharynx or aspiration of draining secretions into the trachea. The term *upper airway cough syndrome* has been coined to encompass the concept that chronic inflammation in the nose and sinuses can cause cough even in the absence of physical drainage into the pharynx. Historical clues suggesting this etiology include a sensation of postnasal drip, frequent throat clearing, and sneezing and rhinorrhea. On speculum examination of the nose, excess mucoid or purulent secretions, inflamed and edematous nasal mucosa, and/or polyps may be seen; in addition, secretions or a cobblestoned appearance of the mucosa along the posterior pharyngeal wall may be noted. Unfortunately, there is no means by which to quantitate postnasal drainage. In many instances, this diagnosis must rely on subjective information provided by the patient. Furthermore, this assessment must also be counterbalanced by the fact that many people who have chronic postnasal drainage do not experience cough.

Linking gastroesophageal reflux to chronic cough poses similar challenges. It is thought that reflux of gastric contents into the lower esophagus may trigger cough via reflex pathways initiated in the esophageal mucosa. Reflux to the level of the pharynx (laryngopharyngeal reflux), with consequent aspiration of gastric contents, causes a chemical bronchitis and possibly pneumonitis that can elicit cough for days afterward, but it is a rare finding among persons with chronic cough. Retrosternal burning after meals or on recumbency, frequent eructation, hoarseness, and throat pain may be indicative of gastroesophageal reflux. Nevertheless, reflux may also elicit minimal or no symptoms. Glottic inflammation detected on laryngoscopy may be a manifestation of recurrent reflux to the level of the throat, but it is a nonspecific finding. Quantification of the frequency and level of reflux requires a somewhat invasive procedure to measure esophageal pH (either nasopharyngeal placement of a catheter with a pH probe into the esophagus for 24 h or endoscopic placement of a radiotransmitter capsule into the esophagus) and, with newer techniques, esophageal pressures (manometry) and nonacid reflux. The precise interpretation of test results that permits an etiologic linking of reflux events and cough remains debated. Again, assigning the cause of cough to gastroesophageal reflux must be weighed against the observation that many people with symptomatic reflux do not experience chronic cough.

Cough alone as a manifestation of asthma is common among children but not among adults. Cough due to asthma in the absence of wheezing

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 38: Cough, Christopher H. Fanta
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

asthma (airflow obstruction on spirometry that varies over time or reverses in response to a bronchodilator) or exclude it with certainty (a negative response to a bronchoprovocation challenge—e.g., with methacholine). In a patient capable of taking reliable measurements, home expiratory peak flow monitoring can be a cost-effective method to support or discount a diagnosis of asthma.

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

reflux do not experience chronic cough.

Cough alone as a manifestation of asthma is common among children but not among adults. Cough due to asthma in the absence of wheezing, shortness of breath, and chest tightness is referred to as "cough-variant asthma." A history suggestive of cough-variant asthma ties the onset of cough to exposure to typical triggers for asthma and the resolution of cough to discontinuation of exposure. Objective testing can establish the diagnosis of asthma (airflow obstruction on spirometry that varies over time or reverses in response to a bronchodilator) or exclude it with certainty (a negative response to a bronchoprovocation challenge—e.g., with methacholine). In a patient capable of taking reliable measurements, home expiratory peak flow monitoring can be a cost-effective method to support or discount a diagnosis of asthma.

Eosinophilic bronchitis causes chronic cough with a normal chest radiograph. This uncommon condition is characterized by sputum eosinophilia in excess of 3% without airflow obstruction or bronchial hyperresponsiveness and is successfully treated with inhaled glucocorticoids. Measurement of an elevated concentration of nitric oxide in exhaled breath has the potential to detect eosinophilic airway inflammation (in asthma or eosinophilic bronchitis) and predict a favorable response to inhaled steroids in persons with chronic cough.

Treatment of chronic cough in a patient with a normal chest radiograph is often empirical and is targeted at the most likely cause(s) of cough as determined by history, physical examination, and possibly pulmonary function testing. Therapy for postnasal drainage depends on the presumed etiology (infection, allergy, or vasomotor rhinitis) and may include systemic antihistamines; decongestants; antibiotics; nasal saline irrigation; and nasal pump sprays with glucocorticoids, antihistamines, or anticholinergics. Antacids histamine type 2 (H_2) receptor antagonists, and proton pump inhibitors are used to neutralize or decrease the production of gastric acid in gastroesophageal reflux disease; dietary changes, elevation of the head and torso during sleep, and medications to improve gastric emptying or impede the flow of refluxate (e.g., alginates) are additional therapeutic measures. Cough-variant asthma typically responds well to inhaled glucocorticoids and intermittent use of inhaled β-agonist bronchodilators.

Patients who fail to respond to treatment targeting the common causes of chronic cough or who have had these causes excluded by appropriate diagnostic testing should, in the opinion of the author, undergo chest CT. Diseases causing cough that may be missed on chest x-ray include tumors, early interstitial lung disease, bronchiectasis, and atypical mycobacterial pulmonary infection. On the other hand, patients with chronic cough who have normal findings on chest examination, lung function testing, oxygenation assessment, and chest CT can be reassured as to the absence of serious pulmonary pathology.

GLOBAL CONSIDERATIONS

Regular exposure to air pollution can cause chronic cough and throat clearing, as well as lower respiratory tract disease. Smoke from cooking and heating fuels in poorly ventilated homes; toxic exposures in work settings lacking implementation of occupational safety standards; and ambient chemicals and particulates in highly polluted outdoor air are all forms of air pollution causing cough. Limited therapeutic options are available; treatment focuses on improving environmental air quality (e.g., use of a stove chimney in the home), removal from the exposure, and use of an appropriate face mask.

In areas of the world where tuberculosis is endemic, chronic cough conjures the possibility of active pulmonary tuberculosis and mandates appropriate evaluation, including chest imaging and sputum analysis.

SYMPTOM-BASED TREATMENT OF COUGH

Empiric treatment of chronic idiopathic cough with inhaled corticosteroids, inhaled anticholinergic bronchodilators, and macrolide antibiotics has been tried without consistent success. Currently available cough suppressants are only modestly effective. Most potent are narcotic cough suppressants, such as codeine, hydrocodone, or morphine, which are thought to act in the "cough center" in the brainstem. The tendency of narcotic cough suppressants to cause drowsiness and constipation and their potential for addictive dependence limit their appeal for long-term use. Dextromethorphan is an over-the-counter, centrally acting cough suppressant with fewer side effects and less efficacy than the narcotic cough suppressants. Dextromethorphan is thought to have a different site of action than narcotic cough suppressants and can be used in combination with them if necessary. Benzonatate is thought to inhibit neural activity of sensory nerves in the cough-reflex pathway. It is generally free of side effects; however, its effectiveness in suppressing cough is variable and unpredictable. Inhaled lidocaine, an inhibitor of voltage-gated sodium channels, provides transient cough suppression, but because of associated oropharyngeal anesthesia, it poses the risk of aspiration.

Attempts to treat cough hypersensitivity syndrome have focused on inhibition of neural pathways. Small case series and randomized clinical trials have indicated benefit from off-label use of gabapentin, pregabalin, or amitriptyline. Recent studies suggest a role for behavioral modification using specialized speech therapy techniques, but widespread application of this modality is currently not practical. Novel cough suppressants without the

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 38: Cough, Christopher H. Fanta
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FURTHER READING

indicated benefit from off-label use of gabapentin, pregabalin, or amitriptyline. Recent studies suggest a role for behavioral modification using specialized speech therapy techniques, but widespread application of this modality is currently not practical. Novel cough suppressants without the limitations of currently available agents are greatly needed. Approaches that are being explored include the development of neurokinin-1 receptor antagonists, transient receptor protein vanilloid-1 (TRPV1) channel antagonists, a promising P2X3 channel antagonist (gefapixant), and novel opioid and opioid-like receptor agonists.

FURTHER READING

Brightling CE et al: Eosinophilic bronchitis as an important cause of chronic cough. Am J Respir Crit Care Med 160:406, 1999. [PubMed: 10430705]

Carroll TL (ed): *Chronic Cough*. San Diego, Plural Publishing, Inc., 2019.

Gibson P et al: Treatment of unexplained chronic cough: CHEST guideline and expert panel report. Chest 149:27, 2016. [PubMed: 26426314]

Kahrilas PJ et al: Chronic cough due to gastroesophageal reflux in adults: CHEST Guideline and Expert Panel Report. Chest 150:1381, 2016.

Morice AH et al: ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. Eur Respir J 55: 1901136, 2020. [PubMed: 31515408]

Ramsay LE et al: Double-blind comparison of losartan, lisinopril and hydrochlorothiazide in hypertensive patients with previous angiotensin converting enzyme inhibitor-associated cough. J Hypertens Suppl 13:S73, 1995. [PubMed: 18800460]

Ryan NM et al: Gabapentin for refractory chronic cough: A randomized, double-blind, placebo-controlled trial. Lancet 380:1583, 2012. [PubMed: 22951084]

Smith JA, Woodcock A: Chronic cough. N Engl J Med 375:1544, 2016. [PubMed: 27797316]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 38: Cough, Christopher H. Fanta
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 39: Hemoptysis

Carolyn M. D'Ambrosio

INTRODUCTION

Hemoptysis is the expectoration of blood from the respiratory tract. Bleeding from the gastrointestinal tract (hematemesis) or nasal cavities (epistaxis) can mimic hemoptysis. Once established as hemoptysis, the degree of blood that is being expectorated (volume and frequency) is the next step as massive or life-threatening hemoptysis (>400 mL of blood in 24 h or >150 mL at one time) requires emergent intervention. This chapter will focus predominantly on non–life-threatening hemoptysis. The source of the bleeding as well as the cause are the next steps when approaching a patient with hemoptysis.

ANATOMY AND PHYSIOLOGY OF HEMOPTYSIS

Hemoptysis can arise from anywhere in the respiratory tract, from the glottis to the alveolus. Most commonly, bleeding arises from the bronchi or medium-sized airways, but a thorough evaluation of the entire respiratory tree is important.

The dual blood supply of the lungs makes it unique. The lungs have both the pulmonary and bronchial circulations. The pulmonary circulation is a low-pressure system that is essential for gas exchange at the alveolar level; in contrast, the bronchial circulation originates from the aorta and, therefore, is a higher-pressure system. The bronchial arteries supply the airways and can neovascularize tumors, dilated airways of bronchiectasis, and cavitary lesions. Most hemoptysis originates from the bronchial circulation, and bleeding from the higher-pressure system makes it more difficult to stop.

ETIOLOGY

Hemoptysis commonly results from infection, malignancy, or vascular disease; however, the differential for bleeding from the respiratory tree is varied and broad. In the United States, the most common causes are viral bronchitis, bronchiectasis, or malignancy. In other parts of the world, infections such as tuberculosis are the most common causes.

Infections

Most blood-tinged sputum and small-volume hemoptysis are due to viral bronchitis. Patients with chronic bronchitis are at risk for bacterial superinfection with organisms such as *Streptococcus pneumoniae, Haemophilus influenzae,* or *Moraxella catarrhalis,* increasing airway inflammation and potential for bleeding. Similarly, patients with bronchiectasis are prone to hemoptysis during exacerbations. Due to recurrent bacterial infection, bronchiectatic airways are dilated, inflamed, and highly vascular, supplied by the bronchial circulation. In several case series, bronchiectasis is the leading cause of massive hemoptysis and subsequent death.

Tuberculosis had long been the most common cause of hemoptysis worldwide, but it is now surpassed in industrialized countries by bronchitis and bronchiectasis. In patients with tuberculosis, development of cavitary disease is frequently the source of bleeding, but rarer complications such as the erosion of a pulmonary artery aneurysm into a preexisting cavity (i.e., Rasmussen's aneurysm) can also be the source.

Other infectious agents such as endemic fungi, *Nocardia*, and nontuberculous mycobacteria can present as cavitary lung disease complicated by hemoptysis. In addition, *Aspergillus* species can develop into mycetomas within preexisting cavities, with neovascularization to these inflamed spaces leading to bleeding. Pulmonary abscesses and necrotizing pneumonia can cause bleeding by devitalizing lung parenchyma. Common responsible organisms include *Staphylococcus aureus, Klebsiella pneumoniae*, and oral anaerobes.

Paragonimiasis can mimic tuberculosis and is another significant cause of hemoptysis seen globally; it is common in Southeast Asia and China,

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 39: Hemoptysis, Carolyn M. D'Ambrosio
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Hemoptysis from a vascular cause can be associated with cardiac disease, pulmonary embolism, arteriovenous malformation, or diffuse alveolar

Access Provided by:
SILVERCHAIR

Paragonimiasis can mimic tuberculosis and is another significant cause of hemoptysis seen globally; it is common in Southeast Asia and China, although cases have been reported in North America from raw crayfish ingestion. It should be considered as a cause of hemoptysis in recent immigrants from endemic areas.

Vascular

Hemoptysis from a vascular cause can be associated with cardiac disease, pulmonary embolism, arteriovenous malformation, or diffuse alveolar hemorrhage (DAH). While the classic description of the sputum expectorated in pulmonary edema (from elevated left end-diastolic pressure) is "pink and frothy," a spectrum of hemoptysis including frank blood can be seen. This observation is particularly true now with the more widespread use of anticoagulants and antiplatelet medications.

Pulmonary embolism with parenchymal infarction can present with hemoptysis, but pulmonary emboli do not commonly cause hemoptysis. An ectatic vessel in an airway or a pulmonary arteriovenous malformation can be a source of bleeding. A rare vascular cause of hemoptysis is the rupture of an aortobronchial fistula; these fistulae arise in the setting of aortic pathology such as aneurysm or pseudoaneurysm and can cause small bleeding episodes that herald massive hemoptysis.

DAH causes significant bleeding into the lung parenchyma but, interestingly, is not often associated with hemoptysis. DAH typically presents with diffuse ground glass opacities on chest imaging. A range of insults cause DAH, including immune-mediated capillaritis from diseases such as systemic lupus erythematosus, toxicity from cocaine and other inhalants, and stem cell transplantation. The so-called "pulmonary-renal" syndromes, including granulomatosis with polyangiitis and anti-glomerular basement membrane disease, may lead to both hemoptysis and hematuria (though one manifestation may be present without the other). A recently identified cause of hemoptysis and DAH is vaping-induced lung injury.

Malignancy

Bronchogenic carcinoma of any histology is a common cause of hemoptysis (both massive and nonmassive). Hemoptysis can indicate airway involvement of the tumor and can be a presenting symptom of carcinoid tumors, vascular lesions that frequently arise in the proximal airways. Small cell and squamous cell carcinomas are frequently central in nature and more likely to erode into major pulmonary vessels, resulting in massive hemoptysis. Pulmonary metastases from distant tumors (e.g., melanoma, sarcoma, adenocarcinomas of the breast and colon) can also cause bleeding. Kaposi's sarcoma, seen in advanced acquired immunodeficiency syndrome, is very vascular and can develop anywhere along the respiratory tract, from the bronchi to the oral cavity.

Mechanical and Other Causes

In addition to infection, vascular disease, and malignancy, other insults to the pulmonary system can cause hemoptysis. Pulmonary endometriosis causes cyclical bleeding known as catamenial hemoptysis. Foreign body aspiration can lead to airway irritation and bleeding. Diagnostic and therapeutic procedures are also potential offenders: pulmonary vein stenosis can result from left atrial procedures, such as pulmonary vein isolation, and pulmonary artery catheters can lead to rupture of the pulmonary artery if the distal balloon is kept inflated. Finally, in the setting of thrombocytopenia, coagulopathy, anticoagulation, or antiplatelet therapy, even minor insults can cause hemoptysis.

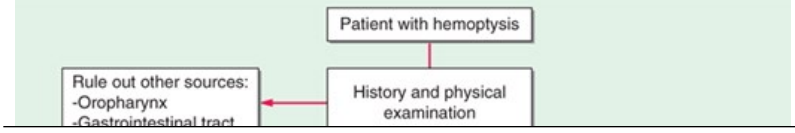
EVALUATION AND MANAGEMENT

History

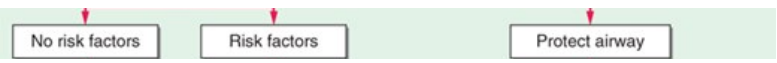
The amount or severity of bleeding is the first step in assessing a patient with hemoptysis. A patient's description of the sputum (e.g., flecks of blood, pink-tinged, or frank blood or clot) is helpful if you cannot examine it. An approach to management of hemoptysis is outlined in **Fig. 39-1**.

FIGURE 39-1

Approach to the management of hemoptysis. CBC, complete blood count; CT, computed tomography; CXR, chest x-ray; UA, urinalysis.

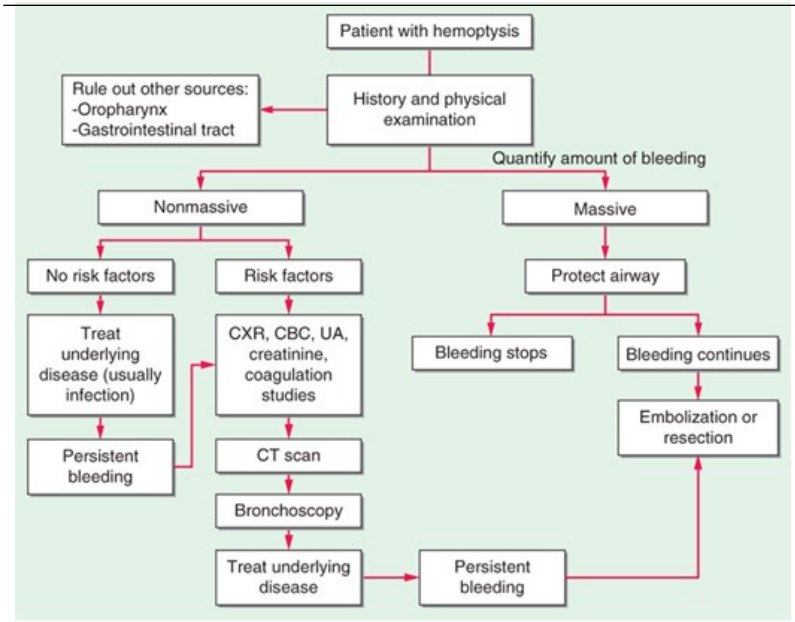


Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 39: Hemoptysis, Carolyn M. D'Ambrosio
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



pink-tinged, or frank blood or clot) is helpful if you cannot examine it. An approach to management of hemoptysis is outlined in Fig. 39-1.

Approach to the management of hemoptysis. CBC, complete blood count; CT, computed tomography; CXR, chest x-ray; UA, urinalysis.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

While there is no agreed-upon volume, blood loss of 400 mL in 24 h or 100–150 mL expectorated at one time should be considered *life-threatening hemoptysis*. These numbers derive from the blood volume of the tracheobronchial tree (generally 100–200 mL). Patients rarely die of exsanguination but, rather, are at risk of death due to asphyxiation from blood filling the airways and airspaces. Most patients cannot describe the volume of their hemoptysis in milliliters, so using referents like cups (one U.S. cup is 236 mL) can be helpful. Fortunately, life-threatening hemoptysis only accounts for 5–15% of cases of hemoptysis.

The history may point to the cause of hemoptysis. Fever, chills, or antecedent cough may suggest infection. A history of smoking or unintentional weight loss makes malignancy more likely. Patients should be asked about inhalational exposures, including vaping. A thorough medical history with careful attention to chronic pulmonary disease should be obtained, with evaluation of risk factors for malignancy and bronchiectatic lung disease (e.g., cystic fibrosis, sarcoidosis).

Physical Examination

Reviewing the vital signs is an important first step. Patients who have life-threatening hemoptysis can have hypoxemia, tachycardia, and hemodynamic instability. As the site of bleeding is important, evaluation of the nasal and oral cavities is imperative. In addition, auscultation of the lungs and seeking other relevant physical findings such as clubbing can point to a cause of the hemoptysis. A focal area of wheezing could suggest a foreign body aspiration. Other signs of a bleeding diathesis (e.g., skin or mucosal ecchymoses and petechiae) or telangiectasias may suggest other etiologies of the hemoptysis.

Diagnostic Studies

Initial studies should include measurement of a complete blood count to assess for infection, anemia, or thrombocytopenia; coagulation parameters; measurement of electrolytes and renal function; and urinalysis to exclude pulmonary-renal disease. Chest imaging is necessary for every patient.

A chest radiograph is usually obtained first, although it frequently does not localize bleeding and can appear normal. In patients without risk factors for malignancy or other abnormalities in the initial evaluation and with a normal chest radiograph, treating for bronchitis and ensuring close follow-up is a reasonable strategy, with further diagnostic workup.

In contrast, patients with risk factors for malignancy (i.e., age >40 or a smoking history) should undergo additional testing. First, chest computed tomography (CT) with contrast should be obtained to better identify masses, bronchiectasis, and parenchymal lesions. A CT looking for pulmonary embolism should be considered if the history and physical examination are consistent with that diagnosis. Following a CT, a flexible bronchoscopy should be performed to exclude bronchogenic carcinoma unless imaging reveals a lesion that can be sampled without bronchoscopy. Small case series show that patients with hemoptysis and unrevealing bronchoscopies have good outcomes.

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 39: Hemoptysis, Carolyn M. D'Ambrosio

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

When the amount of hemoptysis is massive or life-threatening, there are three simultaneous goals: first, protect the nonbleeding lung; second, locate the site of bleeding; and third, control the bleeding.

should be performed to exclude bronchogenic carcinoma unless imaging reveals a lesion that can be sampled without bronchoscopy. Small case series show that patients with hemoptysis and unrevealing bronchoscopies have good outcomes.

Interventions

When the amount of hemoptysis is massive or life-threatening, there are three simultaneous goals: first, protect the nonbleeding lung; second, locate the site of bleeding; and third, control the bleeding.

Protecting the airway and nonbleeding lung is paramount in the management of massive hemoptysis because asphyxiation can happen quickly. If the side of bleeding is known, the patient should be positioned with the bleeding side down to use gravitational advantage to keep blood out of the nonbleeding lung. Endotracheal intubation should be avoided unless truly necessary, since suctioning through an endotracheal tube is a less effective means of removing blood and clot than the cough reflex. If intubation is required, take steps to protect the nonbleeding lung either by selective intubation of one lung (i.e., the nonbleeding lung) or insertion of a double-lumen endotracheal tube.

Locating the bleeding site is sometimes obvious, but frequently, it can be difficult to determine. A chest radiograph, if it shows new opacities, can be helpful in localizing the side or site of bleeding, although this test is not adequate by itself. CT angiography helps by localizing active extravasation. Flexible bronchoscopy may be useful to identify the side of bleeding (although it has only a 50% chance of locating the site). Experts do not agree on the timing of bronchoscopy, although in some cases—cystic fibrosis, for instance—bronchoscopy is *not* recommended because it may delay definitive management. Finally, proceeding directly to angiography is also a reasonable strategy given that it has both diagnostic and therapeutic capabilities.

Controlling the bleeding during an episode of life-threatening hemoptysis can be accomplished in one of three ways: from the airway lumen, from the involved blood vessel, or by surgical resection of both airway and vessel involved. Bronchoscopic measures are generally only temporizing: a flexible bronchoscope can be used to suction clot and insert a balloon catheter or bronchial blocker that occludes the involved airway. Rigid bronchoscopy, done by an interventional pulmonologist or thoracic surgeon, may allow therapeutic interventions of bleeding airway lesions such as photocoagulation and cautery. Because most life-threatening cases of hemoptysis arise from the bronchial circulation, bronchial artery embolization is the procedure of choice for control of the bleeding. However, bronchial artery embolization can have significant complications such as embolization of the anterior spinal artery. However, it is generally successful in the short term, with >80% success rate at controlling bleeding immediately, although bleeding can recur if the underlying disease (e.g., a mycetoma) is not treated. Surgical resection has a high mortality rate (up to 15–40%) and should not be pursued unless initial measures have failed and bleeding is ongoing. Ideal candidates for surgery have localized disease but otherwise normal lung parenchyma.

ACKNOWLEDGMENT

Anna K. Brady and Patricia A. Kritek contributed to this chapter in the 20th edition, and some material from that chapter has been retained here.

FURTHER READING

Adelman M et al: Cryptogenic hemoptysis: Clinical features, bronchoscopic findings, and natural history in 67 patients. Ann Intern Med 102:829, 1985. [PubMed: 3994194]

Flume PA et al: CF pulmonary guidelines. Pulmonary complications: Hemoptysis and pneumothorax. Am J Respir Crit Care Med 182:298, 2010. [PubMed: 20675678]

Hirshberg B et al: Hemoptysis: Etiology, evaluation, and outcome in a tertiary care hospital. Chest 112:440, 1997. [PubMed: 9266882]

Jean-baptiste E: Clinical assessment and management of massive hemoptysis. Crit Care Med 28:1642, 2000. [PubMed: 10834728]

Johnson JL: Manifestations of hemoptysis: How to manage minor, moderate, and massive bleeding. Postgrad Med 112:4:101, 2002. [PubMed: 12400152]

Layden JE et al: Pulmonary illness related to e-cigarette, reply. N Engl J Med 382:903, 2020. [PubMed: 31491072]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 39: Hemoptysis, Carolyn M. D'Ambrosio
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Sopko DR, Smith TP: Bronchial artery embolization for massive hemoptysis. Semin Intervent Radiol 28:48, 2011. [PubMed: 22379276]

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

Lordan JL et al: The pulmonary physician in critical care: Illustrative case 7. Assessment and management of massive hemoptysis. Thorax 58:814, 2003. [PubMed: 12947147]

Sopko DR, Smith TP: Bronchial artery embolization for massive hemoptysis. Semin Intervent Radiol 28:48, 2011. [PubMed: 22379276]

Downloaded 2022-1-3 5:47 P Your IP is 141.216.78.40
Chapter 39: Hemoptysis, Carolyn M. D'Ambrosio
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 40: Hypoxia and Cyanosis

Joseph Loscalzo

HYPOXIA

The fundamental purpose of the cardiorespiratory system is to deliver O_2 and nutrients to cells and to remove CO_2 and other metabolic products from them. Proper maintenance of this function depends not only on intact cardiovascular and respiratory systems, but also on an adequate number of red blood cells and hemoglobin and a supply of inspired gas containing adequate O_2.

RESPONSES TO HYPOXIA

Decreased O_2 availability to cells typically results in an inhibition of oxidative phosphorylation and increased anaerobic glycolysis. This switch from aerobic to anaerobic metabolism, the Pasteur effect, reduces the rate of adenosine 5′-triphosphate (ATP) production. In severe hypoxia, when ATP production is inadequate to meet the energy requirements of ionic and osmotic equilibrium, cell membrane depolarization leads to uncontrolled Ca^{2+} influx and activation of Ca^{2+}-dependent phospholipases and proteases. These events, in turn, cause cell swelling, activation of apoptotic pathways, and, ultimately, cell death.

The adaptations to hypoxia are mediated, in part, by the upregulation of genes encoding a variety of proteins, including glycolytic enzymes, such as phosphoglycerate kinase and phosphofructokinase, as well as the glucose transporters Glut-1 and Glut-2; and by growth factors, such as vascular endothelial growth factor (VEGF) and erythropoietin, which enhance erythrocyte production. The hypoxia-induced increase in expression of these and other key proteins is governed by the hypoxia-sensitive transcription factor, hypoxia-inducible factor-1 (HIF-1).

During hypoxia, systemic arterioles dilate, at least in part, by opening of K_{ATP} channels in vascular smooth-muscle cells due to the hypoxia-induced reduction in ATP concentration. By contrast, in pulmonary vascular smooth-muscle cells, inhibition of K^+ channels causes depolarization, which, in turn, activates voltage-gated Ca^{2+} channels, raising the cytosolic $[Ca^{2+}]$ and causing smooth-muscle cell contraction. Hypoxia-induced pulmonary arterial constriction shunts blood away from poorly ventilated portions toward better ventilated portions of the lung (i.e., improves ventilation-perfusion mismatch); however, it also increases pulmonary vascular resistance and right ventricular afterload.

Effects on the Central Nervous System

Changes in the central nervous system (CNS), particularly the higher centers, are especially important consequences of hypoxia. Acute hypoxia causes impaired judgment, motor incoordination, and a clinical picture resembling acute alcohol intoxication. High-altitude illness is characterized by headache secondary to cerebral vasodilation, gastrointestinal symptoms, dizziness, insomnia, fatigue, or somnolence. Pulmonary arterial and sometimes venous constriction causes capillary leakage and high-altitude pulmonary edema (HAPE) **(Chap. 37)**, which intensifies hypoxia, further promoting vasoconstriction. Rarely, high-altitude cerebral edema (HACE) develops, which is manifest by severe headache and papilledema and can cause coma. As hypoxia becomes more severe, the regulatory centers of the brainstem are affected, and death usually results from respiratory failure.

Effects on the Cardiovascular System

Acute hypoxia stimulates the chemoreceptor reflex arc to induce venoconstriction and systemic arterial vasodilation. These acute changes are accompanied by transiently increased myocardial contractility, which is followed by depressed myocardial contractility with prolonged hypoxia.

CAUSES OF HYPOXIA

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 40: Hypoxia and Cyanosis, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

curve **(see Fig. 98-2)** is displaced to the right, with greater quantities of O_2 released at any level of tissue Po_2. Arterial hypoxemia, that is, a reduction of O_2 saturation of arterial blood (Sao_2), and consequent cyanosis are likely to be more marked when such depression of Pao_2 results from pulmonary

Respiratory Hypoxia

When hypoxia occurs from respiratory failure, Pa_{O_2} declines, and when respiratory failure is persistent, the hemoglobin-oxygen (Hb-O_2) dissociation curve **(see Fig. 98-2)** is displaced to the right, with greater quantities of O_2 released at any level of tissue P_{O_2}. Arterial hypoxemia, that is, a reduction of O_2 saturation of arterial blood (Sa_{O_2}), and consequent cyanosis are likely to be more marked when such depression of Pa_{O_2} results from pulmonary disease than when the depression occurs as the result of a decline in the fraction of oxygen in inspired air (F_{IO_2}). In this latter situation, Pa_{CO_2} falls secondary to anoxia-induced hyperventilation and the Hb-O_2 dissociation curve is displaced to the left, limiting the decline in Sa_{O_2} at any level of Pa_{O_2}.

The most common cause of respiratory hypoxia is *ventilation-perfusion mismatch* resulting from perfusion of poorly ventilated alveoli. Respiratory hypoxemia may also be caused by *hypoventilation*, in which case it is associated with an elevation of Pa_{CO_2} **(Chap. 285)**. These two forms of respiratory hypoxia are usually correctable by inspiring 100% O_2 for several minutes. A third cause of respiratory hypoxia is shunting of blood across the lung from the pulmonary arterial to the venous bed (*intrapulmonary right-to-left shunting*) by perfusion of nonventilated portions of the lung, as in pulmonary atelectasis or through pulmonary arteriovenous connections. The low Pa_{O_2} in this situation is only partially corrected by an F_{IO_2} of 100%.

Hypoxia Secondary to High Altitude

As one ascends rapidly to 3000 m (~10,000 ft), the reduction of the O_2 content of inspired air (F_{IO_2}) leads to a decrease in alveolar P_{O_2} to ~60 mmHg, and a condition termed *high-altitude illness* develops (see above). At higher altitudes, arterial saturation declines rapidly and symptoms become more serious; and at 5000 m, unacclimated individuals usually cease to be able to function normally owing to the changes in CNS function described above.

Hypoxia Secondary to Right-to-Left Extrapulmonary Shunting

From a physiologic viewpoint, this cause of hypoxia resembles intrapulmonary right-to-left shunting but is caused by congenital cardiac malformations, such as tetralogy of Fallot, transposition of the great arteries, atrial or ventricular septal defect, patent ductus arteriosus, and Eisenmenger's syndrome **(Chap. 269)**. As in pulmonary right-to-left shunting, the Pa_{O_2} cannot be restored to normal with inspiration of 100% O_2.

Anemic Hypoxia

A reduction in hemoglobin concentration of the blood is accompanied by a corresponding decline in the O_2-carrying capacity of the blood. Although the Pa_{O_2} is normal in anemic hypoxia, the absolute quantity of O_2 transported per unit volume of blood is diminished. As the anemic blood passes through the capillaries and the usual quantity of O_2 is removed from it, the P_{O_2} and saturation in the venous blood decline to a greater extent than normal.

Carbon Monoxide (CO) Intoxication

(See also Chap. 463) Hemoglobin that binds with CO (carboxy-hemoglobin [COHb]) is unavailable for O_2 transport. In addition, the presence of COHb shifts the Hb-O_2 dissociation curve to the left **(see Fig. 98-2)** so that O_2 is unloaded only at lower tensions, further contributing to tissue hypoxia.

Circulatory Hypoxia

As in anemic hypoxia, the Pa_{O_2} is usually normal, but venous and tissue P_{O_2} values are reduced as a consequence of reduced tissue perfusion and greater tissue O_2 extraction. This pathophysiology leads to an increased arterial-mixed venous O_2 difference (a-v-O_2 difference), or gradient. Generalized circulatory hypoxia occurs in heart failure **(Chap. 257)** and in most forms of shock **(Chap. 303)**.

Specific Organ Hypoxia

Localized circulatory hypoxia may occur as a result of decreased perfusion secondary to arterial obstruction, as in localized atherosclerosis in any

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 40: Hypoxia and Cyanosis, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

adequate perfusion to more vital organs in patients with reduced cardiac output secondary to heart failure or hypovolemic shock, vasoconstriction may reduce perfusion in the limbs and skin, causing hypoxia of these regions.

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

Localized circulatory hypoxia may occur as a result of decreased perfusion secondary to arterial obstruction, as in localized atherosclerosis in any vascular bed, or as a consequence of vasoconstriction, as observed in Raynaud's phenomenon (Chap. 281). Localized hypoxia may also result from venous obstruction and the resultant expansion of interstitial fluid causing arteriolar compression and, thereby, reduction of arterial inflow. Edema, which increases the distance through which O_2 must diffuse before it reaches cells, can also cause localized hypoxia. In an attempt to maintain adequate perfusion to more vital organs in patients with reduced cardiac output secondary to heart failure or hypovolemic shock, vasoconstriction may reduce perfusion in the limbs and skin, causing hypoxia of these regions.

Increased O_2 Requirements

If the O_2 consumption of tissues is elevated without a corresponding increase in perfusion, tissue hypoxia ensues and the Po_2 in venous blood declines. Ordinarily, the clinical picture of patients with hypoxia due to an elevated metabolic rate, as in fever or thyrotoxicosis, is quite different from that in other types of hypoxia: the skin is warm and flushed owing to increased cutaneous blood flow that dissipates the excessive heat produced, and cyanosis is usually absent.

Exercise is a classic example of increased tissue O_2 requirements. These increased demands are normally met by several mechanisms operating simultaneously: (1) increase in the cardiac output and ventilation and, thus, O_2 delivery to the tissues; (2) a preferential shift in blood flow to the exercising muscles by changing vascular resistances in the circulatory beds of exercising tissues, directly and/or reflexly; (3) an increase in O_2 extraction from the delivered blood and a widening of the arteriovenous O_2 difference; and (4) a reduction in the pH of the tissues and capillary blood, shifting the Hb-O_2 curve to the right **(see Fig. 98-2)**, and unloading more O_2 from hemoglobin. If the capacity of these mechanisms is exceeded, then hypoxia, especially of the exercising muscles, will result.

Improper Oxygen Utilization

Cyanide (Chap. 459) and several other similarly acting poisons cause cellular hypoxia by impairing electron transport in mitochondria, thereby limiting oxidative phosphorylation and ATP production. The tissues are unable to use O_2, and as a consequence, the venous blood tends to have a high O_2 tension. This condition has been termed *histotoxic hypoxia*.

ADAPTATION TO HYPOXIA

An important component of the respiratory response to hypoxia originates in special chemosensitive cells in the carotid and aortic bodies and in the respiratory center in the brainstem. The stimulation of these cells by hypoxia increases ventilation, with a loss of CO_2, and can lead to respiratory alkalosis. When combined with the metabolic acidosis resulting from the production of lactic acid, the serum bicarbonate level declines (Chap. 55).

With the reduction of Pao_2, cerebrovascular resistance decreases and cerebral blood flow increases in an attempt to maintain O_2 delivery to the brain. However, when the reduction of Pao_2 is accompanied by hyperventilation and a reduction of $Paco_2$, cerebrovascular resistance rises, cerebral blood flow falls, and tissue hypoxia intensifies.

The diffuse, systemic vasodilation that occurs in generalized hypoxia increases the cardiac output. In patients with underlying heart disease, the requirements of peripheral tissues for an increase of cardiac output with hypoxia may precipitate congestive heart failure. In patients with ischemic heart disease, a reduced Pao_2 may intensify myocardial ischemia and further impair left ventricular function.

One of the important compensatory mechanisms for chronic hypoxia is an increase in the hemoglobin concentration and in the number of red blood cells in the circulating blood, that is, the development of polycythemia induced by erythropoietin production (Chap. 103). In persons with chronic hypoxemia secondary to prolonged residence at a high altitude (>13,000 ft, 4200 m), a condition termed *chronic mountain sickness* develops. This disorder is characterized by a blunted respiratory drive, reduced ventilation, erythrocytosis, cyanosis, weakness, right ventricular enlargement secondary to pulmonary hypertension, and even stupor.

CYANOSIS

Cyanosis refers to a bluish color of the skin and mucous membranes resulting from an increased quantity of reduced hemoglobin (i.e., deoxygenated

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 40: Hypoxia and Cyanosis, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

cherry-colored flush, rather than cyanosis, is caused by COHb (Chap. 459).

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Cyanosis refers to a bluish color of the skin and mucous membranes resulting from an increased quantity of reduced hemoglobin (i.e., deoxygenated hemoglobin) or of hemoglobin derivatives (e.g., methemoglobin or sulfhemoglobin) in the small blood vessels of those tissues. It is usually most marked in the lips, nail beds, ears, and malar eminences. Cyanosis, especially if developed recently, is more commonly detected by a family member than the patient. The florid skin characteristic of polycythemia vera **(Chap. 103)** must be distinguished from the true cyanosis discussed here. A cherry-colored flush, rather than cyanosis, is caused by COHb **(Chap. 459)**.

The degree of cyanosis is modified by the color of the cutaneous pigment and the thickness of the skin, as well as by the state of the cutaneous capillaries. The accurate clinical detection of the presence and degree of cyanosis is difficult, as proved by oximetric studies. In some instances, central cyanosis can be detected reliably when the Sa_{O_2} has fallen to 85%; in others, particularly in dark-skinned persons, it may not be detected until it has declined to 75%. In the latter case, examination of the mucous membranes in the oral cavity and the conjunctivae rather than examination of the skin is more helpful in the detection of cyanosis.

The increase in the quantity of reduced hemoglobin in the mucocutaneous vessels that produces cyanosis may be brought about either by an increase in the quantity of venous blood as a result of dilation of the venules (including precapillary venules) or by a reduction in the Sa_{O_2} in the capillary blood. In general, cyanosis becomes apparent when the concentration of reduced hemoglobin in capillary blood exceeds 40 g/L (4 g/dL).

It is the *absolute*, rather than the *relative*, quantity of reduced hemoglobin that is important in producing cyanosis. Thus, in a patient with severe anemia, the *relative* quantity of reduced hemoglobin in the venous blood may be very large when considered in relation to the total quantity of hemoglobin in the blood. However, since the concentration of the latter is markedly reduced, the *absolute* quantity of reduced hemoglobin may still be low, and, therefore, patients with severe anemia and even *marked* arterial desaturation may not display cyanosis. Conversely, the higher the total hemoglobin content, the greater is the tendency toward cyanosis; thus, patients with marked polycythemia tend to be cyanotic at higher levels of Sa_{O_2} than patients with normal hematocrit values. Likewise, local passive congestion, which causes an increase in the total quantity of reduced hemoglobin in the vessels in a given area, may cause cyanosis. Cyanosis is also observed when nonfunctional hemoglobin, such as methemoglobin (consequential or acquired) or sulfhemoglobin **(Chap. 98)**, is present in blood.

Cyanosis may be subdivided into central and peripheral types. In *central* cyanosis, the Sa_{O_2} is reduced or an abnormal hemoglobin derivative is present, and the mucous membranes and skin are both affected. *Peripheral* cyanosis is due to a slowing of blood flow and abnormally great extraction of O_2 from normally saturated arterial blood; it results from vasoconstriction and diminished peripheral blood flow, such as occurs in cold exposure, shock, congestive failure, and peripheral vascular disease. Often in these conditions, the mucous membranes of the oral cavity, including the sublingual mucosa, may be spared. Clinical differentiation between central and peripheral cyanosis may not always be straightforward, and in conditions such as cardiogenic shock with pulmonary edema, there may be a mixture of both types.

DIFFERENTIAL DIAGNOSIS

Central Cyanosis

Decreased Sa_{O_2} results from a marked reduction in the Pa_{O_2} **(Table 40-1)**. This reduction may be brought about by a decline in the F_{IO_2} without sufficient compensatory alveolar hyperventilation to maintain alveolar P_{O_2}. Cyanosis usually becomes manifest in an ascent to an altitude of 4000 m (13,000 ft).

TABLE 40-1

Causes of Cyanosis

Central Cyanosis
Decreased arterial oxygen saturation
Decreased atmospheric pressure—high altitude
Impaired pulmonary function
Inhomogeneity in pulmonary ventilation and perfusion (perfusion of hypoventilated alveoli)

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 40: Hypoxia and Cyanosis, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

(15,000 ft).

TABLE 40-1

Causes of Cyanosis

Central Cyanosis
Decreased arterial oxygen saturation
Decreased atmospheric pressure—high altitude
Impaired pulmonary function
Alveolar hypoventilation
Inhomogeneity in pulmonary ventilation and perfusion (perfusion of hypoventilated alveoli)
Impaired oxygen diffusion
Anatomic shunts
Certain types of congenital heart disease
Pulmonary arteriovenous fistulas
Multiple small intrapulmonary shunts
Hemoglobin with low affinity for oxygen
Hemoglobin abnormalities
Methemoglobinemia—hereditary, acquired
Sulfhemoglobinemia—acquired
Carboxyhemoglobinemia (not true cyanosis)
Peripheral Cyanosis
Reduced cardiac output
Cold exposure
Redistribution of blood flow from extremities
Arterial obstruction
Venous obstruction

Seriously *impaired pulmonary function*, through perfusion of unventilated or poorly ventilated areas of the lung or alveolar hypoventilation, is a common cause of central cyanosis **(Chap. 285)**. This condition may occur acutely, as in extensive pneumonia or pulmonary edema, or chronically, with chronic pulmonary diseases (e.g., emphysema). In the latter situation, secondary polycythemia is generally present and clubbing of the fingers (see below) may occur. Another cause of reduced Sa_{O_2} is *shunting of systemic venous blood into the arterial circuit*. Certain forms of congenital heart

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 40: Hypoxia and Cyanosis, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

by these fistulae depends on their size and number. They occur with some frequency in hereditary hemorrhagic telangiectasia. Sa_{O_2} reduction and

with chronic pulmonary diseases (e.g., emphysema). In the latter situation, secondary polycythemia is generally present and clubbing of the fingers (see below) may occur. Another cause of reduced Sao_2 is *shunting of systemic venous blood into the arterial circuit*. Certain forms of congenital heart disease are associated with cyanosis on this basis (see above and **Chap. 269**).

Pulmonary arteriovenous fistulae may be congenital or acquired, solitary or multiple, and microscopic or massive. The severity of cyanosis produced by these fistulae depends on their size and number. They occur with some frequency in hereditary hemorrhagic telangiectasia. Sao_2 reduction and cyanosis may also occur in some patients with cirrhosis, presumably as a consequence of pulmonary arteriovenous fistulae or portal vein–pulmonary vein anastomoses.

In patients with cardiac or pulmonary right-to-left shunts, the presence and severity of cyanosis depend on the size of the shunt relative to the systemic flow and on the Hb-O_2 saturation of the venous blood. With increased extraction of O_2 from the blood by the exercising muscles, the venous blood returning to the right side of the heart is more unsaturated than at rest, and shunting of this blood intensifies the cyanosis. Secondary polycythemia occurs frequently in patients in this setting and contributes to the cyanosis.

Cyanosis can be caused by small quantities of circulating methemoglobin (Hb Fe^{3+}) and by even smaller quantities of sulfhemoglobin **(Chap. 98)**; both of these hemoglobin derivatives impair oxygen delivery to the tissues. Although they are uncommon causes of cyanosis, these abnormal hemoglobin species should be sought by spectroscopy when cyanosis is not readily explained by malfunction of the circulatory or respiratory systems. Generally, digital clubbing does not occur with them.

Peripheral Cyanosis

Probably the most common cause of peripheral cyanosis is the normal vasoconstriction resulting from exposure to cold air or water. When cardiac output is reduced, cutaneous vasoconstriction occurs as a compensatory mechanism so that blood is diverted from the skin to more vital areas such as the CNS and heart, and cyanosis of the extremities may result even though the arterial blood is normally saturated.

Arterial obstruction to an extremity, as with an embolus, or arteriolar constriction, as in cold-induced vasospasm (Raynaud's phenomenon) **(Chap. 281)**, generally results in pallor and coldness, and there may be associated cyanosis. Venous obstruction, as in thrombophlebitis or deep venous thrombosis, dilates the subpapillary venous plexuses and thereby intensifies cyanosis.

APPROACH TO THE PATIENT

Cyanosis

Certain features are important in arriving at the cause of cyanosis:

1. It is important to ascertain the time of onset of cyanosis. Cyanosis present since birth or infancy is usually due to congenital heart disease.
2. Central and peripheral cyanosis must be differentiated. Evidence of disorders of the respiratory or cardiovascular systems is helpful. Massage or gentle warming of a cyanotic extremity will increase peripheral blood flow and abolish peripheral, but not central, cyanosis.
3. The presence or absence of clubbing of the digits (see below) should be ascertained. The combination of cyanosis and clubbing is frequent in patients with congenital heart disease and right-to-left shunting and is seen occasionally in patients with pulmonary disease, such as lung abscess or pulmonary arteriovenous fistulae. In contrast, peripheral cyanosis or acutely developing central cyanosis is *not* associated with clubbed digits.
4. Pao_2 and Sao_2 should be determined, and in patients with cyanosis in whom the mechanism is obscure, spectroscopic examination of the blood should be performed to look for abnormal types of hemoglobin (critical in the differential diagnosis of cyanosis).

CLUBBING

The selective bulbous enlargement of the distal segments of the fingers and toes due to proliferation of connective tissue, particularly on the dorsal surface, is termed *clubbing*; there is also increased sponginess of the soft tissue at the base of the clubbed nail. Clubbing may be hereditary, idiopathic, instances, it is occupational, for example, in jackhammer operators.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 40: Hypoxia and Cyanosis, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

The selective bulbous enlargement of the distal segments of the fingers and toes due to proliferation of connective tissue, particularly on the dorsal surface, is termed *clubbing*; there is also increased sponginess of the soft tissue at the base of the clubbed nail. Clubbing may be hereditary, idiopathic, or acquired and associated with a variety of disorders, including cyanotic congenital heart disease (see above), infective endocarditis, and a variety of pulmonary conditions (among them primary and metastatic lung cancer, bronchiectasis, asbestosis, sarcoidosis, lung abscess, cystic fibrosis, tuberculosis, and mesothelioma), as well as with some gastrointestinal diseases (including inflammatory bowel disease and hepatic cirrhosis). In some instances, it is occupational, for example, in jackhammer operators.

Clubbing in patients with primary and metastatic lung cancer, mesothelioma, bronchiectasis, or hepatic cirrhosis may be associated with *hypertrophic osteoarthropathy*. In this condition, the subperiosteal formation of new bone in the distal diaphyses of the long bones of the extremities causes pain and symmetric arthritis-like changes in the shoulders, knees, ankles, wrists, and elbows. The diagnosis of hypertrophic osteoarthropathy may be confirmed by bone radiograph or magnetic resonance imaging (MRI). Although the mechanism of clubbing is unclear, it appears to be secondary to humoral substances that cause dilation of the vessels of the distal digits as well as growth factors released from platelet precursors in the digital circulation. In certain circumstances, clubbing is reversible, such as following lung transplantation for cystic fibrosis.

FURTHER READING

Callemeyn J et al: Clubbing and hypertrophic osteoarthropathy: Insights into diagnosis, pathophysiology, and clinical significance. Acta Clin Belg 22:1, 2016.

MacIntyre NR: Tissue hypoxia: Implications for the respiratory clinician. Respir Care 59:1590, 2014. [PubMed: 25161296]

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 40: Hypoxia and Cyanosis, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 41: Edema

Joseph Loscalzo

PLASMA AND INTERSTITIAL FLUID EXCHANGE

Approximately two-thirds of total body water is intracellular and one-third is extracellular. One-fourth of the latter is in the plasma, and the remainder comprises the interstitial fluid. Edema represents an excess of interstitial fluid that has become evident clinically.

There is constant interchange of fluid between the two compartments of the extracellular fluid. The hydrostatic pressure within the capillaries and the colloid oncotic pressure in the interstitial fluid promote the movement of water and diffusible solutes from plasma to the interstitium. This movement is most prominent at the arterial origin of the capillary and falls progressively with the decline in intracapillary pressure and the rise in oncotic pressure toward the venular end. Fluid is returned from the interstitial space into the vascular system largely through the lymphatic system. These interchanges of fluids are normally balanced so that the volumes of the intravascular and interstitial compartments remain constant. However, a net movement of fluid from the intravascular to the interstitial spaces takes place and may be responsible for the development of edema under the following conditions: (1) an increase in intracapillary hydrostatic pressure; (2) inadequate lymphatic drainage; (3) reductions in the oncotic pressure in the plasma; (4) damage to the capillary endothelial barrier; and (5) increases in the oncotic pressure in the interstitial space.

REDUCTION OF EFFECTIVE ARTERIAL VOLUME

In many forms of edema, the effective arterial blood volume, a parameter that represents the filling of the arterial tree and that effectively perfuses the tissues, is reduced. Underfilling of the arterial tree may be caused by a reduction of cardiac output and/or systemic vascular resistance, by the pooling of blood in the splanchnic veins (as in cirrhosis), and by hypoalbuminemia **(Fig. 41-1A)**. As a consequence of this underfilling, a series of physiologic responses designed to restore the effective arterial volume to normal are set into motion. A key element of these responses is the renal retention of sodium and, therefore, water, thereby restoring effective arterial volume, but sometimes also leading to the development or intensification of edema.

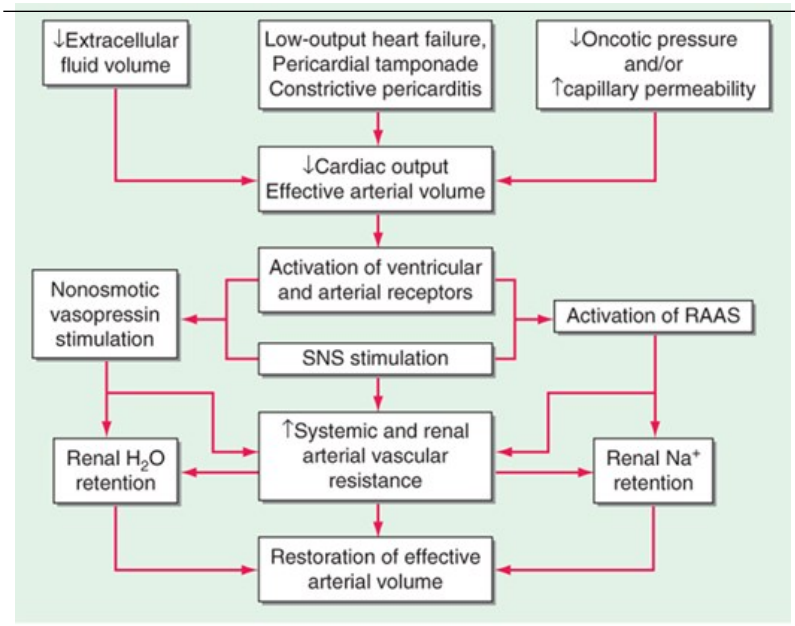
FIGURE 41-1

Clinical conditions in which a decrease in cardiac output (*A*) and systemic vascular resistance (*B*) cause arterial underfilling with resulting neurohumoral activation and renal sodium and water retention. In addition to activating the neurohumoral axis, adrenergic stimulation causes renal vasoconstriction and enhances sodium and fluid transport by the proximal tubule epithelium. AVP, arginine vasopressin; RAAS, renin-angiotensin aldosterone system; SNS, sympathetic nervous system. *(From Annals of Internal Medicine, RW Schrie, Body fluid volume regulation in health and disease: a unifying hypothesis,113(2),155–159, Copyright © 1990, American College of Physicians. All Rights Reserved. Reprinted with the permission of American College of Physicians, Inc.)*

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 41: Edema, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

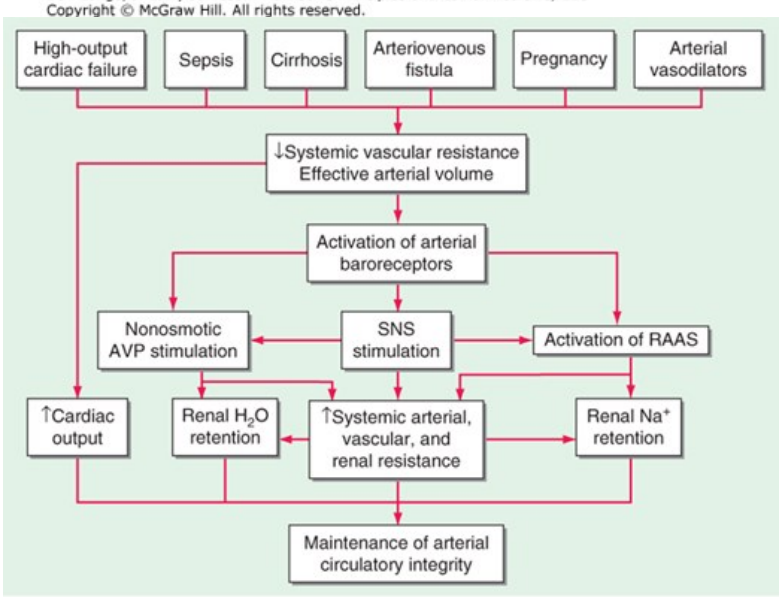
ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS



A

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.



B

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

RENAL FACTORS AND THE RENIN-ANGIOTENSIN-ALDOSTERONE SYSTEM

The diminished renal blood flow characteristic of states in which the effective arterial blood volume is reduced is translated by the renal juxtaglomerular cells (specialized myoepithelial cells surrounding the afferent arteriole) into a signal for increased renin release. Renin is an enzyme with a molecular mass of about 40,000 Da that acts on its substrate, angiotensinogen, an α_2-globulin synthesized by the liver, to release angiotensin I, a decapeptide, which in turn is converted to angiotensin II (AII), an octapeptide. AII has generalized vasoconstrictor properties, particularly on the renal efferent arterioles. This action reduces the hydrostatic pressure in the peritubular capillaries, whereas the increased filtration fraction raises the colloid osmotic pressure in these vessels, thereby enhancing salt and water reabsorption in the proximal tubule as well as in the ascending limb of the loop of Henle.

The renin-angiotensin-aldosterone system (RAAS) operates as both a hormonal and paracrine system. Its activation causes sodium and water

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 41: Edema, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

favoring edema formation. Blockade of the action of aldosterone by spironolactone or eplerenone (aldosterone antagonists) or by amiloride (a blocker of epithelial sodium channels) often induces a moderate diuresis in edematous states.

The renin-angiotensin-aldosterone system (RAAS) operates as both a hormonal and paracrine system. Its activation causes sodium and water retention and thereby contributes to edema formation. Blockade of the conversion of angiotensin I to AII and blockade of the AII receptors enhance sodium and water excretion and reduce many forms of edema. AII that enters the systemic circulation stimulates the production of aldosterone by the zona glomerulosa of the adrenal cortex. Aldosterone in turn enhances sodium reabsorption (and potassium excretion) by the collecting tubule, further favoring edema formation. Blockade of the action of aldosterone by spironolactone or eplerenone (aldosterone antagonists) or by amiloride (a blocker of epithelial sodium channels) often induces a moderate diuresis in edematous states.

ARGININE VASOPRESSIN

(See also Chap. 381) The secretion of arginine vasopressin (AVP) by the posterior pituitary gland occurs in response to increased intracellular osmolar concentration; by stimulating V_2 receptors, AVP increases the reabsorption of free water in the distal tubules and collecting ducts of the kidneys, thereby increasing total-body water. Circulating AVP is elevated in many patients with heart failure secondary to a nonosmotic stimulus associated with decreased effective arterial volume and reduced compliance of the left atrium. Such patients fail to show the normal reduction of AVP with a reduction of osmolality, contributing to edema formation and hyponatremia.

ENDOTHELIN-1

This potent peptide vasoconstrictor is released by endothelial cells. Its concentration in the plasma is elevated in patients with severe heart failure and contributes to renal vasoconstriction, sodium retention, and edema.

NATRIURETIC PEPTIDES

Atrial distention causes release into the circulation of atrial natriuretic peptide (ANP), a polypeptide. A high-molecular-weight precursor of ANP is stored in secretory granules within atrial myocytes. A closely related natriuretic peptide (pre-prohormone brain natriuretic peptide [BNP]) is stored primarily in ventricular myocytes and is released when ventricular diastolic pressure rises. Released ANP and BNP (which is derived from its precursor) bind to the natriuretic receptor-A, which causes (1) excretion of sodium and water by augmenting glomerular filtration rate, inhibiting sodium reabsorption in the proximal tubule, and inhibiting release of renin and aldosterone; and (2) dilation of arterioles and venules by antagonizing the vasoconstrictor actions of AII, AVP, and sympathetic stimulation. Thus, elevated levels of natriuretic peptides have the capacity to oppose sodium retention in hypervolemic and edematous states.

Although circulating levels of ANP and BNP are elevated in heart failure and in cirrhosis with ascites, these natriuretic peptides are not sufficiently potent to prevent edema formation. Indeed, in edematous states, resistance to the actions of natriuretic peptides may be increased, further reducing their effectiveness.

Further discussion of the control of sodium and water balance is found in **Chap. S1**.

CLINICAL CAUSES OF EDEMA

A weight gain of several kilograms usually precedes overt manifestations of generalized edema. *Anasarca* refers to gross, generalized edema. *Ascites* **(Chap. 50)** and *hydrothorax* refer to accumulation of excess fluid in the peritoneal and pleural cavities, respectively, and are considered special forms of edema.

Edema is recognized by the persistence of an indentation of the skin after pressure known as “pitting” edema. In its more subtle form, edema may be detected by noting that after the stethoscope is removed from the chest wall, the rim of the bell leaves an indentation on the skin of the chest for a few minutes. Edema may be present when the ring on a finger fits more snugly than in the past or when a patient complains of difficulty putting on shoes, particularly in the evening. Edema may also be recognized by puffiness of the face, which is most readily apparent in the periorbital areas owing to relative tissue laxity.

GENERALIZED EDEMA

The differences among the major causes of generalized edema are shown in **Table 41-1**. Cardiac, renal, hepatic, or nutritional disorders are responsible for a large majority of patients with generalized edema. Consequently, the differential diagnosis of generalized edema should be directed toward identifying or excluding these several conditions.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 41: Edema, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

responsible for a large majority of patients with generalized edema. Consequently, the differential diagnosis of generalized edema should be directed toward identifying or excluding these several conditions.

TABLE 41-1

Principal Causes of Generalized Edema: History, Physical Examination, and Laboratory Findings

ORGAN SYSTEM	HISTORY	PHYSICAL EXAMINATION	LABORATORY FINDINGS
Cardiac	Dyspnea with exertion prominent—often associated with orthopnea—or paroxysmal nocturnal dyspnea	Elevated jugular venous pressure, ventricular (S_3) gallop; occasionally with displaced or dyskinetic apical pulse; peripheral cyanosis, cool extremities, small pulse pressure when severe	Elevated urea nitrogen-to-creatinine ratio common; serum sodium often diminished; elevated natriuretic peptides
Hepatic	Dyspnea uncommon, except if associated with significant degree of ascites; most often a history of ethanol abuse	Frequently associated with ascites; jugular venous pressure normal or low; blood pressure lower than in renal or cardiac disease; one or more additional signs of chronic liver disease (jaundice, palmar erythema, Dupuytren's contracture, spider angiomata, male gynecomastia; asterixis and other signs of encephalopathy) may be present	If severe, reductions in serum albumin, cholesterol, other hepatic proteins (transferrin, fibrinogen); liver enzymes elevated, depending on the cause and acuity of liver injury; tendency toward hypokalemia, respiratory alkalosis; macrocytosis from folate deficiency
Renal (CRF)	Usually chronic: may be associated with uremic signs and symptoms, including decreased appetite, altered (metallic or fishy) taste, altered sleep pattern, difficulty concentrating, restless legs, or myoclonus; dyspnea can be present, but generally less prominent than in heart failure	Elevated blood pressure; hypertensive retinopathy; nitrogenous fetor; pericardial friction rub in advanced cases with uremia	Elevation of serum creatinine and cystatin C; albuminuria; hyperkalemia, metabolic acidosis, hyperphosphatemia, hypocalcemia, anemia (usually normocytic)
Renal (NS)	Childhood diabetes mellitus; plasma cell dyscrasias	Periorbital edema; hypertension	Proteinuria (≥3.5 g/d); hypoalbuminemia; hypercholesterolemia; microscopic hematuria

Abbreviations: CRF, chronic renal failure; NS, nephrotic syndrome.

Source: Reproduced with permission from GM Chertow, in E Braunwald, L Goldman (eds): Approach to the patient with edema, in Primary Cardiology, 2nd ed. pg: 117–128. Philadelphia, Saunders, 2003.

Heart Failure

(See also Chap. 257) In heart failure, the impaired systolic emptying of the ventricle(s) and/or the impairment of ventricular relaxation promotes an accumulation of blood in the venous circulation at the expense of the effective arterial volume. In addition, the activation of the sympathetic nervous system and the RAAS (see above) acts in concert to cause renal vasoconstriction and reduction of glomerular filtration and salt and water retention. Sodium and water retention continue, and the increment in blood volume accumulates in the venous circulation, raising venous and intracapillary pressure and resulting in edema (Fig. 41-1).

The presence of overt cardiac disease, as manifested by cardiac enlargement and/or ventricular hypertrophy, together with clinical evidence of cardiac failure, such as dyspnea, basilar rales, venous distention, and hepatomegaly, usually indicates that edema results from heart failure. Noninvasive tests

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 41: Edema, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Edema of Renal Disease

(See also Chap. 314) The edema that occurs during the acute phase of glomerulonephritis is characteristically associated with hematuria

failure, such as dyspnea, basilar rales, venous distention, and hepatomegaly, usually indicates that edema results from heart failure. Noninvasive tests such as electrocardiography, echocardiography, and measurements of BNP (or N-terminal proBNP [NT-proBNP]) are helpful in establishing the diagnosis of heart disease. The edema of heart failure typically occurs in the dependent portions of the body.

Edema of Renal Disease

(See also Chap. 314) The edema that occurs during the acute phase of glomerulonephritis is characteristically associated with hematuria, proteinuria, and hypertension. In most instances, the edema results from primary retention of sodium and water by the kidneys owing to renal dysfunction. This state differs from most forms of heart failure in that it is characterized by a normal (or sometimes even increased) cardiac output. Patients with *chronic* renal failure may also develop edema due to primary renal retention of sodium and water.

Nephrotic Syndrome and Other Hypoalbuminemic States

The primary alteration in the nephrotic syndrome is a diminished colloid oncotic pressure due to losses of large quantities (≥3.5 g/d) of protein into the urine and hypoalbuminemia (<3.0 g/dL). As a result of the reduced colloid osmotic pressure, the sodium and water that are retained cannot be confined within the vascular compartment, and total and effective arterial blood volumes decline. This process initiates the edema-forming sequence of events described above, including activation of the RAAS. The nephrotic syndrome may occur during the course of a variety of kidney diseases, including glomerulonephritis, diabetic glomerulosclerosis, and hypersensitivity reactions. The edema is diffuse, symmetric, and most prominent in the dependent areas; periorbital edema is most prominent in the morning.

Hepatic Cirrhosis

(See also Chap. 344) This condition is characterized, in part, by hepatic venous outflow obstruction, which in turn expands the splanchnic blood volume, and hepatic lymph formation. Intrahepatic hypertension acts as a stimulus for renal sodium retention and causes a reduction of effective arterial blood volume. These alterations are frequently complicated by hypoalbuminemia secondary to reduced hepatic synthesis, as well as peripheral arterial vasodilation. These effects reduce the effective arterial blood volume, leading to activation of the sodium- and water-retaining mechanisms described above **(Fig. 41-1B)**. The concentration of circulating aldosterone often is elevated by the failure of the liver to metabolize this hormone. Initially, the excess interstitial fluid is localized preferentially proximal (upstream) to the congested portal venous system, causing ascites **(Chap. 50)**. In later stages, particularly when there is severe hypoalbuminemia, peripheral edema may develop. A sizable accumulation of ascitic fluid may increase intraabdominal pressure and impede venous return from the lower extremities and contribute to the accumulation of the edema.

Drug-Induced Edema

A large number of widely used drugs can cause edema **(Table 41-2)**. Mechanisms include renal vasoconstriction (nonsteroidal anti-inflammatory drugs and cyclosporine), arteriolar dilation (vasodilators), augmented renal sodium reabsorption (steroid hormones), and capillary damage.

TABLE 41-2

Drugs Associated with Edema Formation

Nonsteroidal anti-inflammatory drugs
Antihypertensive agents
Direct arterial/arteriolar vasodilators
Hydralazine
Clonidine
Methyldopa
Guanethidine
α-Adrenergic antagonists

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 41: Edema, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

drugs and cyclosporine), arteriolar dilation (vasodilators), augmented renal sodium reabsorption (steroid hormones), and capillary damage.

TABLE 41-2

Drugs Associated with Edema Formation

Nonsteroidal anti-inflammatory drugs
Antihypertensive agents
Direct arterial/arteriolar vasodilators
Hydralazine
Clonidine
Methyldopa
Guanethidine
Minoxidil
Calcium channel antagonists
α-Adrenergic antagonists
Thiazolidinediones
Steroid hormones
Glucocorticoids
Anabolic steroids
Estrogens
Progestins
Cyclosporine
Growth hormone
Immunotherapies
Interleukin 2
OKT3 monoclonal antibody

Source: Reproduced with permission from GM Chertow, in E Braunwald, L Goldman (eds): Approach to the patient with edema, in Primary Cardiology, 2nd ed. pg: 117–128. Philadelphia, Saunders, 2003.

Edema of Nutritional Origin

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 41: Edema, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

develops or becomes intensified when famished subjects are first provided with an adequate diet. The ingestion of more food may increase the quantity of sodium ingested, which is then retained along with water. So-called refeeding edema also may be linked to increased release of insulin,

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

A diet grossly deficient in calories and particularly in protein over a prolonged period may produce hypoproteinemia and edema. The latter may be intensified by the development of beriberi heart disease, which also is of nutritional origin, in which multiple peripheral arteriovenous fistulae result in reduced effective systemic perfusion and effective arterial blood volume, thereby enhancing edema formation **(Chap. 333)** (Fig. 41-1B). Edema develops or becomes intensified when famished subjects are first provided with an adequate diet. The ingestion of more food may increase the quantity of sodium ingested, which is then retained along with water. So-called refeeding edema also may be linked to increased release of insulin, which directly increases tubular sodium reabsorption. In addition to hypoalbuminemia, hypokalemia and caloric deficits may be involved in the edema of starvation.

LOCALIZED EDEMA

In thrombophlebitis, varicose veins, and primary venous valve failure, the hydrostatic pressure in the capillary bed upstream (proximal) of the obstruction increases so that an abnormal quantity of fluid is transferred from the vascular to the interstitial space, which may give rise to localized edema. The latter may also occur in lymphatic obstruction caused by chronic lymphangitis, resection of regional lymph nodes, filariasis, and genetic (frequently called primary) lymphedema. The latter is particularly intractable because restriction of lymphatic flow results in both an increase in intracapillary pressure and increased protein concentration in the interstitial fluid, which act in concert to aggravate fluid retention.

Other Causes of Edema

These causes include hypothyroidism (myxedema) due to deposition of hyaluronic acid; hyperthyroidism (pretibial myxedema secondary to Graves' disease), in which edema is typically nonpitting and, in Graves' disease, exogenous hypercortisolism; pregnancy; and administration of estrogens and vasodilators, particularly dihydropyridines such as nifedipine.

DISTRIBUTION OF EDEMA

The distribution of edema is an important guide to its cause. Edema associated with heart failure tends to be more extensive in the legs and to be accentuated in the evening, a feature also determined largely by posture. When patients with heart failure are confined to bed, edema may be most prominent in the presacral region.

Edema resulting from hypoproteinemia, as occurs in the nephrotic syndrome, characteristically is generalized, but it is especially evident in the very soft tissues of the eyelids and face and tends to be most pronounced in the morning owing to the recumbent posture assumed during the night. Less common causes of facial edema include trichinosis, allergic reactions, and myxedema. Edema limited to one leg or to one or both arms is usually the result of venous and/or lymphatic obstruction. Unilateral paralysis reduces lymphatic and venous drainage on the affected side and may also be responsible for unilateral edema. In patients with obstruction of the superior vena cava, edema is confined to the face, neck, and upper extremities in which the venous pressure is elevated compared with that in the lower extremities.

APPROACH TO THE PATIENT

Edema

An important first question is whether the edema is localized or generalized. If it is localized, the local phenomena that may be responsible should be identified. If the edema is generalized, one should determine if there is serious hypoalbuminemia, e.g., serum albumin <3.0 g/dL. If so, the history, physical examination, urinalysis, and other laboratory data will help evaluate the question of cirrhosis, severe malnutrition, or the nephrotic syndrome as the underlying disorder. If hypoalbuminemia is not present, it should be determined if there is evidence of heart failure severe enough to promote generalized edema. Finally, it should be ascertained as to whether or not the patient has an adequate urine output or if there is significant oliguria or anuria. **These abnormalities are discussed in Chaps. 52, 310, and 311.**

FURTHER READING

Clark AL, Cleland JG: Causes and treatment of oedema in patients with heart failure. Nature Rev Cardiol 10:156, 2013.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 41: Edema, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Ferrell RE et al: GJC2 missense mutations cause human lymphedema. Am J Hum Genet 86:943, 2010. [PubMed: 20537300]

Damman K et al: Congestion in chronic systolic heart failure is related to renal dysfunction and increased mortality. Eur J Heart Fail 12:974, 2010. [PubMed: 20685688]

Ferrell RE et al: *GJC2* missense mutations cause human lymphedema. Am J Hum Genet 86:943, 2010. [PubMed: 20537300]

Frison S et al: Omitting edema measurement: How much acute malnutrition are we missing? Am J Clin Nutr 102:1176, 2015. [PubMed: 26377162]

Levick JR, Michel CC: Microvascular fluid exchange and the revised Starling principle. Cardiovascular Res 87:198, 2010.

Telinius N, Hjortdal VE: Role of the lymphatic vasculature in cardiovascular medicine. Heart 105:1777, 2019. [PubMed: 31585946]

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 41: Edema, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 43: Palpitations

Joseph Loscalzo

INTRODUCTION

Palpitations are extremely common among patients who present to their internists and can best be defined as a “thumping,” “pounding,” or “fluttering” sensation in the chest. This sensation can be either intermittent or sustained and either regular or irregular. Most patients interpret palpitations as an unusual awareness of the heartbeat and become especially concerned when they sense that they have had “skipped” or “missing” heartbeats. Palpitations are often noted when the patient is quietly resting, during which time other stimuli are minimal. Palpitations that are positional generally reflect a structural process within (e.g., atrial myxoma) or adjacent to (e.g., mediastinal mass) the heart.

Palpitations are brought about by cardiac (43%), psychiatric (31%), miscellaneous (10%), and unknown (16%) causes, according to one large series. Among the cardiovascular causes are premature atrial and ventricular contractions, supraventricular and ventricular arrhythmias, mitral valve prolapse (with or without associated arrhythmias), aortic insufficiency, atrial myxoma, myocarditis, and pulmonary embolism. Intermittent palpitations are commonly caused by premature atrial or ventricular contractions: the post-extrasystolic beat is sensed by the patient owing to the increase in ventricular end-diastolic dimension following the pause in the cardiac cycle and the increased strength of contraction (post-extrasystolic potentiation) of that beat. Regular, sustained palpitations can be caused by regular supraventricular and ventricular tachycardias. Irregular, sustained palpitations can be caused by atrial fibrillation. It is important to note that most arrhythmias are not associated with palpitations. In those that are, it is often useful either to ask the patient to “tap out” the rhythm of the palpitations or to take his or her pulse during palpitations. In general, hyperdynamic cardiovascular states caused by catecholaminergic stimulation from exercise, stress, or pheochromocytoma can lead to palpitations. Palpitations are common among athletes, especially older endurance athletes. In addition, the enlarged ventricle of aortic regurgitation and accompanying hyperdynamic precordium frequently lead to the sensation of palpitations. Other factors that enhance the strength of myocardial contraction, including tobacco, caffeine, aminophylline, atropine, thyroxine, cocaine, and amphetamines, can cause palpitations.

Psychiatric causes of palpitations include panic attacks or disorders, anxiety states, and somatization, alone or in combination. Patients with psychiatric causes for palpitations more commonly report a longer duration of the sensation (>15 min) and other accompanying symptoms than do patients with other causes. Among the miscellaneous causes of palpitations are thyrotoxicosis, drugs (see above) and ethanol, spontaneous skeletal muscle contractions of the chest wall, pheochromocytoma, and systemic mastocytosis.

APPROACH TO THE PATIENT

Palpitations

The principal goal in assessing patients with palpitations is to determine whether the symptom is caused by a life-threatening arrhythmia. Patients with preexisting coronary artery disease (CAD) or risk factors for CAD are at greatest risk for ventricular arrhythmias **(Chap. 246)** as a cause for palpitations. In addition, the association of palpitations with other symptoms suggesting hemodynamic compromise, including syncope or lightheadedness, supports this diagnosis. Palpitations caused by sustained tachyarrhythmias in patients with CAD can be accompanied by angina pectoris or dyspnea, and, in patients with ventricular dysfunction (systolic or diastolic), aortic stenosis, hypertrophic cardiomyopathy, or mitral stenosis (with or without CAD), can be accompanied by dyspnea from increased left atrial and pulmonary venous pressure.

Key features of the physical examination that will help confirm or refute the presence of an arrhythmia as a cause for palpitations (as well as its adverse hemodynamic consequences) include measurement of the vital signs, assessment of the jugular venous pressure and pulse, and auscultation of the chest and precordium. A resting electrocardiogram can be used to document the arrhythmia. If exertion is known to induce the arrhythmia and accompanying palpitations, exercise electrocardiography can be used to make the diagnosis. If the arrhythmia is sufficiently infrequent, other methods must be used, including continuous electrocardiographic (Holter) monitoring; telephonic monitoring, through which the assessment of patients with (infrequent) recurrent, unexplained palpitations. The use of a diary or an electronic marker to indicate the timing of palpitations sensed by the patient is essential for appropriate interpretation of these studies.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 43: Palpitations, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

APPROACH TO THE PATIENT

Palpitations

The principal goal in assessing patients with palpitations is to determine whether the symptom is caused by a life-threatening arrhythmia. Patients with preexisting coronary artery disease (CAD) or risk factors for CAD are at greatest risk for ventricular arrhythmias **(Chap. 246)** as a cause for palpitations. In addition, the association of palpitations with other symptoms suggesting hemodynamic compromise, including syncope or lightheadedness, supports this diagnosis. Palpitations caused by sustained tachyarrhythmias in patients with CAD can be accompanied by angina pectoris or dyspnea, and, in patients with ventricular dysfunction (systolic or diastolic), aortic stenosis, hypertrophic cardiomyopathy, or mitral stenosis (with or without CAD), can be accompanied by dyspnea from increased left atrial and pulmonary venous pressure.

Key features of the physical examination that will help confirm or refute the presence of an arrhythmia as a cause for palpitations (as well as its adverse hemodynamic consequences) include measurement of the vital signs, assessment of the jugular venous pressure and pulse, and auscultation of the chest and precordium. A resting electrocardiogram can be used to document the arrhythmia. If exertion is known to induce the arrhythmia and accompanying palpitations, exercise electrocardiography can be used to make the diagnosis. If the arrhythmia is sufficiently infrequent, other methods must be used, including continuous electrocardiographic (Holter) monitoring; telephonic monitoring, through which the patient can transmit an electrocardiographic tracing during a sensed episode; loop recordings (external or implantable), which can capture the electrocardiographic event for later review; and mobile (self-monitoring) cardiac outpatient telemetry. Data suggest that Holter monitoring is of limited clinical utility, while the implantable loop recorder and mobile cardiac outpatient telemetry are safe and possibly more cost-effective in the assessment of patients with (infrequent) recurrent, unexplained palpitations. The use of a diary or an electronic marker to indicate the timing of palpitations sensed by the patient is essential for appropriate interpretation of these studies.

Most patients with palpitations do not have serious arrhythmias or underlying structural heart disease. If sufficiently troubling to the patient, occasional benign atrial or ventricular premature contractions can often be managed with beta-blocker therapy. Palpitations incited by alcohol, tobacco, or illicit drugs need to be managed by abstention, while those caused by pharmacologic agents should be addressed by considering alternative therapies when appropriate or possible. Psychiatric causes of palpitations may benefit from cognitive therapy or pharmacotherapy. The physician should note that palpitations are at the very least bothersome and, on occasion, frightening to the patient. Once serious causes for the symptom have been excluded, the patient should be reassured that the palpitations will not adversely affect prognosis.

FURTHER READING

Crossland S, Berkin L: Problem based review: The patient with palpitations. Acute Med 11:169, 2012. [PubMed: 22993750]

Jamshed N et al: Emergency management of palpitations in the elderly: Epidemiology, diagnostic approaches, and therapeutic options. Clin Geriatr Med 29:205, 2013. [PubMed: 23177608]

Martson HR et al: Mobile self-monitoring ECG devices to diagnose arrhythmias that coincide with palpitations: A scoping review. Healthcare (Basel) 7:pii: E96, 2019.

Sakh R et al: Insertable cardiac monitors: current indications and devices. Expert Rev Med Devices 16:45, 2019. [PubMed: 30522350]

Weber BE, Kapoor WN: Evaluation and outcomes of patients with palpitations. Am J Med 100:138, 1996. [PubMed: 8629647]

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 43: Palpitations, Joseph Loscalzo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 44: Dysphagia

Ikuo Hirano; Peter J. Kahrilas

INTRODUCTION

Dysphagia—difficulty with swallowing—refers to problems with the transit of food or liquid from the mouth to the hypopharynx or through the esophagus. Severe dysphagia can compromise nutrition, cause aspiration, and reduce quality of life. Additional terminology pertaining to swallowing dysfunction is as follows. *Aphagia* (inability to swallow) typically denotes complete esophageal obstruction, most commonly encountered in the acute setting of a food bolus or foreign body impaction. *Odynophagia* refers to painful swallowing, typically resulting from mucosal ulceration within the oropharynx or esophagus. It commonly is accompanied by dysphagia, but the converse is not true. *Globus pharyngeus* is a foreign body sensation localized in the neck that does not interfere with swallowing and sometimes is relieved by swallowing. *Transfer dysphagia* frequently results in nasal regurgitation or pulmonary aspiration during swallowing and is characteristic of oropharyngeal dysphagia. *Phagophobia* (fear of swallowing) and *refusal to swallow* may be psychogenic or related to anticipatory anxiety about food bolus obstruction, odynophagia, or aspiration.

PHYSIOLOGY OF SWALLOWING

Swallowing begins with a voluntary (oral) phase that includes preparation during which food is masticated and mixed with saliva. This is followed by a transfer phase during which the bolus is pushed into the pharynx by the tongue. Bolus entry into the hypopharynx initiates the pharyngeal swallow response, which is centrally mediated and involves a complex series of actions, the net result of which is to propel food through the pharynx into the esophagus while preventing its entry into the airway. To accomplish this, the larynx is elevated and pulled forward, actions that also facilitate upper esophageal sphincter (UES) opening. Tongue pulsion then propels the bolus through the UES, followed by a peristaltic contraction that clears residue from the pharynx and through the esophagus. The lower esophageal sphincter (LES) relaxes as the food enters the esophagus and remains relaxed until the peristaltic contraction has delivered the bolus into the stomach. Peristaltic contractions elicited in response to a swallow are called *primary peristalsis* and involve sequenced inhibition followed by contraction of the musculature along the entire length of the esophagus. The inhibition that precedes the peristaltic contraction is called *deglutitive inhibition*. Local distention of the esophagus anywhere along its length, as may occur with gastroesophageal reflux, activates *secondary peristalsis* that begins at the point of distention and proceeds distally. Tertiary esophageal contractions are nonperistaltic, disordered esophageal contractions that may be observed to occur spontaneously during fluoroscopic observation.

The musculature of the oral cavity, pharynx, UES, and cervical esophagus is striated and directly innervated by lower motor neurons carried in cranial nerves **(Fig. 44-1)**. Oral cavity muscles are innervated by the fifth (trigeminal) and seventh (facial) cranial nerves; the tongue, by the twelfth (hypoglossal) cranial nerve. Pharyngeal muscles are innervated by the ninth (glossopharyngeal) and tenth (vagus) cranial nerves.

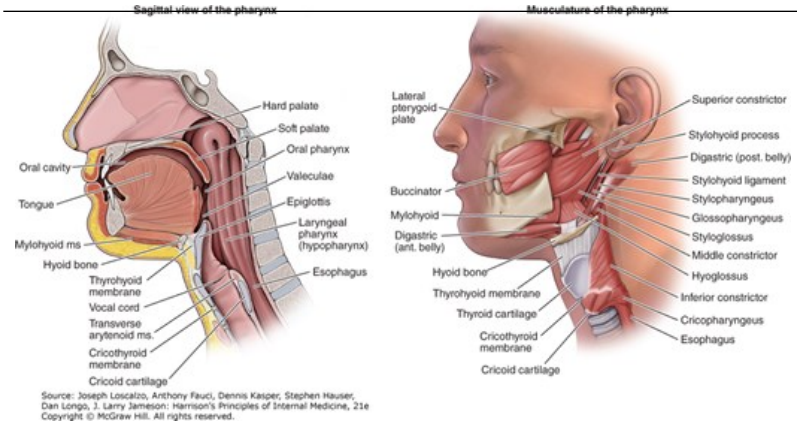
FIGURE 44-1

Sagittal and diagrammatic views of the musculature involved in enacting oropharyngeal swallowing. Note the dominance of the tongue in the sagittal view and the intimate relationship between the entrance to the larynx (airway) and the esophagus. In the resting configuration illustrated, the esophageal inlet is closed. This is transiently reconfigured such that the esophageal inlet is open and the laryngeal inlet closed during swallowing. *(Adapted from PJ Kahrilas, in DW Gelfand and JE Richter [eds]: Dysphagia: Diagnosis and Treatment. New York, Igaku-Shoin Medical Publishers, 1989, pp. 11–28.)*

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 44: Dysphagia, Ikuo Hirano; Peter J. Kahrilas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Physiologically, the UES consists of the cricopharyngeus muscle, the adjacent inferior pharyngeal constrictor, and the proximal portion of the cervical esophagus. UES innervation is derived from the vagus nerve, whereas the innervation to the musculature acting on the UES to facilitate its opening during swallowing comes from the fifth, seventh, and twelfth cranial nerves. The UES remains closed at rest owing to both its inherent elastic properties and neurogenically mediated contraction of the cricopharyngeus muscle. UES opening during swallowing involves both cessation of vagal excitation to the cricopharyngeus and simultaneous contraction of the suprahyoid and geniohyoid muscles that pull open the UES in conjunction with the upward and forward displacement of the larynx.

The neuromuscular apparatus for peristalsis is distinct in proximal and distal parts of the esophagus. The cervical esophagus, like the pharyngeal musculature, consists of striated muscle and is directly innervated by lower motor neurons of the vagus nerve. Peristalsis in the proximal esophagus is governed by the sequential activation of the vagal motor neurons in the nucleus ambiguus. In contrast, the distal esophagus and LES are composed of smooth muscle and are controlled by excitatory and inhibitory neurons within the esophageal myenteric plexus. Medullary preganglionic neurons from the dorsal motor nucleus of the vagus trigger peristalsis via these ganglionic neurons during primary peristalsis. Neurotransmitters of the excitatory ganglionic neurons are acetylcholine and substance P; those of the inhibitory neurons are vasoactive intestinal peptide and nitric oxide. Peristalsis results from the patterned activation of inhibitory followed by excitatory ganglionic neurons, with progressive dominance of the inhibitory neurons distally. Similarly, LES relaxation occurs with the onset of deglutitive inhibition and persists until the peristaltic sequence is complete. At rest, the LES is contracted because of excitatory ganglionic stimulation and its intrinsic myogenic tone, a property that distinguishes it from the adjacent esophagus. The function of the LES is supplemented by the surrounding muscle of the right diaphragmatic crus, which acts as an external sphincter during inspiration, cough, or abdominal straining.

PATHOPHYSIOLOGY OF DYSPHAGIA

Dysphagia can be subclassified both by location and by the circumstances in which it occurs. With respect to location, distinct considerations apply to oral, pharyngeal, or esophageal dysphagia. Normal transport of an ingested bolus depends on the consistency and size of the bolus, the caliber of the lumen, the integrity of peristaltic contraction, and deglutitive inhibition of both the UES and the LES. Dysphagia caused by an oversized bolus or a narrow lumen is called *structural dysphagia*, whereas dysphagia due to abnormalities of peristalsis or impaired sphincter relaxation after swallowing is called *propulsive* or *motor dysphagia*. More than one mechanism may be operative in a patient with dysphagia. Scleroderma commonly presents with absent peristalsis as well as a weakened LES that predisposes patients to peptic stricture formation. Likewise, radiation therapy for head and neck cancer may compound the functional deficits in the oropharyngeal swallow attributable to the tumor and cause cervical esophageal stenosis. It is worth noting that in addition to bolus transit, symptom reporting of dysphagia is dependent upon intact sensory innervation and central nervous system perception.

Oral and Pharyngeal (Oropharyngeal) Dysphagia

Oral-phase dysphagia is associated with poor bolus formation and control so that food has prolonged retention within the oral cavity and may seep out of the mouth. Drooling and difficulty in initiating swallowing are other characteristic signs. Poor bolus control also may lead to premature spillage of food into the hypopharynx with resultant aspiration into the trachea or regurgitation into the nasal cavity. Pharyngeal-phase dysphagia is associated with retention of food in the pharynx due to poor tongue or pharyngeal propulsion or obstruction at the UES. Signs and symptoms of concomitant hoarseness or cranial nerve dysfunction may be associated with oropharyngeal dysphagia.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 44: Dysphagia, Ikuo Hirano; Peter J. Kahrilas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

aspiration and malnutrition. Medullary nuclei directly innervate the oropharynx. Lateralization of pharyngeal dysphagia implies either a structural pharyngeal lesion or a neurologic process that selectively targeted the ipsilateral brainstem nuclei or cranial nerve. Advances in functional brain

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

Oropharyngeal dysphagia may be due to neurologic, muscular, structural, iatrogenic, infectious, and metabolic causes. Iatrogenic, neurologic, and structural pathologies are most common. Iatrogenic causes include surgery and radiation, often in the setting of head and neck cancer. Neurogenic dysphagia resulting from cerebrovascular accidents, Parkinson's disease, and amyotrophic lateral sclerosis is a major source of morbidity related to aspiration and malnutrition. Medullary nuclei directly innervate the oropharynx. Lateralization of pharyngeal dysphagia implies either a structural pharyngeal lesion or a neurologic process that selectively targeted the ipsilateral brainstem nuclei or cranial nerve. Advances in functional brain imaging have elucidated an important role of the cerebral cortex in swallow function and dysphagia. Asymmetry in the cortical representation of the pharynx provides an explanation for the dysphagia that occurs as a consequence of unilateral cortical cerebrovascular accidents.

Oropharyngeal structural lesions causing dysphagia include Zenker's diverticulum, cricopharyngeal bar, and neoplasia. Zenker's diverticulum typically is encountered in elderly patients. In addition to dysphagia, patients may present with regurgitation of particulate food debris, aspiration, and halitosis. The pathogenesis is related to stenosis of the cricopharyngeus that causes diminished opening of the UES and results in increased hypopharyngeal pressure during swallowing with development of a pulsion diverticulum immediately above the cricopharyngeus in a region of potential weakness known as Killian's dehiscence. A cricopharyngeal bar, appearing as a prominent indentation behind the lower third of the cricoid cartilage, is related to Zenker's diverticulum in that it involves limited distensibility of the cricopharyngeus and can lead to the formation of a Zenker's diverticulum. However, a cricopharyngeal bar is a common radiographic finding, and most patients with transient cricopharyngeal bars are asymptomatic, making it important to rule out alternative etiologies of dysphagia before treatment. Furthermore, cricopharyngeal bars may be secondary to other neuromuscular disorders that impair opening of the UES.

Since the pharyngeal phase of swallowing occurs in less than a second, rapid-sequence fluoroscopy is necessary to evaluate for functional abnormalities. Adequate fluoroscopic examination requires that the patient be conscious and cooperative. The study incorporates recordings of swallow sequences during ingestion of food and liquids of varying consistencies. The pharynx is examined to detect bolus retention, regurgitation into the nose, or aspiration into the trachea. Timing and integrity of pharyngeal contraction and opening of the UES with a swallow are analyzed to assess both aspiration risk and the potential for swallow therapy. Structural abnormalities of the oropharynx, especially those that may require biopsies, also should be assessed by direct laryngoscopic examination.

Esophageal Dysphagia

The adult esophagus measures 18–26 cm in length and is anatomically divided into the cervical esophagus, extending from the pharyngoesophageal junction to the suprasternal notch, and the thoracic esophagus, which continues to the diaphragmatic hiatus. When distended, the esophageal lumen has internal dimensions of about 2 cm in the anteroposterior plane and 3 cm in the lateral plane. Solid food dysphagia becomes common when the lumen is narrowed to <13 mm, but also can occur with larger diameters in the setting of poorly masticated food or motor dysfunction. Circumferential lesions are more likely to cause dysphagia than are lesions that involve only a partial circumference of the esophageal wall. The most common structural causes of dysphagia are Schatzki's rings, eosinophilic esophagitis, and peptic strictures. Dysphagia also occurs in the setting of gastroesophageal reflux disease without a stricture, perhaps on the basis of altered esophageal sensation, reduced esophageal mural distensibility, or motor dysfunction.

Propulsive disorders leading to esophageal dysphagia result from abnormalities of peristalsis and/or deglutitive inhibition, potentially affecting the cervical or thoracic esophagus. Since striated muscle pathology usually involves both the oropharynx and the cervical esophagus, the clinical manifestations usually are dominated by oropharyngeal dysphagia. Diseases affecting smooth muscle involve both the thoracic esophagus and the LES. A dominant manifestation of this, absent peristalsis, refers to either the complete absence of swallow-induced contraction (absent contractility) or the presence of nonperistaltic, disordered contractions. Absent peristalsis and failure of deglutitive LES relaxation are the defining features of achalasia. In diffuse esophageal spasm (DES), LES function is normal, with the disordered motility restricted to the esophageal body. Absent contractility combined with severe weakness of the LES is a pattern commonly found in patients with scleroderma.

APPROACH TO THE PATIENT

Dysphagia

Figure 44-2 shows an algorithm for the approach to a patient with dysphagia.

HISTORY

The patient history is extremely valuable in making a presumptive diagnosis or at least substantially limiting the differential diagnoses in most

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 44: Dysphagia, Ikuo Hirano; Peter J. Kahrilas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Nasal regurgitation and tracheobronchial aspiration manifest by coughing with swallowing are hallmarks of oropharyngeal dysphagia. Severe cough with swallowing may also be a sign of a tracheoesophageal fistula. The presence of hoarseness may be another important diagnostic clue

HISTORY

The patient history is extremely valuable in making a presumptive diagnosis or at least substantially limiting the differential diagnoses in most patients. Key elements of the history are the localization of dysphagia, the circumstances in which dysphagia is experienced, other symptoms associated with dysphagia, and progression. Dysphagia that localizes to the suprasternal notch may indicate either an oropharyngeal or an esophageal etiology as distal dysphagia is referred proximally about 30% of the time. Dysphagia that localizes to the chest is esophageal in origin. Nasal regurgitation and tracheobronchial aspiration manifest by coughing with swallowing are hallmarks of oropharyngeal dysphagia. Severe cough with swallowing may also be a sign of a tracheoesophageal fistula. The presence of hoarseness may be another important diagnostic clue. When hoarseness precedes dysphagia, the primary lesion is usually laryngeal; hoarseness that occurs after the development of dysphagia may result from compromise of the recurrent laryngeal nerve by a malignancy. The type of food causing dysphagia is an important consideration. Intermittent dysphagia that occurs only with solid food implies structural dysphagia, whereas constant dysphagia with both liquids and solids strongly suggests an esophageal motor abnormality. Two caveats to this pattern are that despite having a motor abnormality, patients with scleroderma generally develop mild dysphagia for solids only and that patients with oropharyngeal dysphagia often have greater difficulty managing liquids than solids. Dysphagia that is progressive over the course of weeks to months raises concern for neoplasia. Episodic dysphagia to solids that is unchanged or slowly progressive over years indicates a benign disease process such as a Schatzki ring or eosinophilic esophagitis. Food impaction with a prolonged inability to pass an ingested bolus even with ingestion of liquid is typical of a structural dysphagia. Chest pain may accompany dysphagia whether it is related to motor disorders, structural disorders, or reflux disease. A prolonged history of heartburn preceding the onset of dysphagia is suggestive of peptic stricture and, infrequently, esophageal adenocarcinoma. A history of prolonged nasogastric intubation, esophageal or head and neck surgery, ingestion of caustic agents or pills, previous radiation or chemotherapy, or associated mucocutaneous diseases may help isolate the cause of dysphagia. With accompanying odynophagia, which usually is indicative of ulceration, infectious or pill-induced esophagitis should be suspected. In patients with AIDS or other immunocompromised states, esophagitis due to opportunistic infections such as *Candida*, herpes simplex virus, or cytomegalovirus and to tumors such as Kaposi's sarcoma and lymphoma should be considered. A history of atopy increases concerns for eosinophilic esophagitis, which is most prevalent in Caucasian male patients between the ages of 20 and 40 years. Medication use should identify agents associated with pill esophagitis and narcotics that are associated with opioid-induced esophageal dysmotility.

PHYSICAL EXAMINATION

Physical examination is important in the evaluation of oral and pharyngeal dysphagia because dysphagia is usually only one of many manifestations of a more global disease process. Signs of bulbar or pseudobulbar palsy, including dysarthria, dysphonia, ptosis, and tongue atrophy, in addition to evidence of generalized neuromuscular disease, should be elicited. The neck should be examined for thyromegaly or lymphadenopathy. A careful inspection of the mouth and pharynx should disclose inflammatory or infectious lesions. Missing dentition can interfere with mastication and exacerbate an existing cause of dysphagia. Physical examination is less helpful in the evaluation of esophageal dysphagia as most relevant pathology is restricted to the esophagus. The notable exception is skin disease. Changes in the skin and oral mucosa may suggest a diagnosis of scleroderma or mucocutaneous diseases such as pemphigoid, lichen planus, and epidermolysis bullosa, all of which can involve the esophagus.

DIAGNOSTIC PROCEDURES

Although most instances of dysphagia are attributable to benign disease processes, dysphagia is also a cardinal symptom of several malignancies, making it an important symptom to evaluate. Cancer may result in dysphagia most commonly as the result of intraluminal obstruction (esophageal or proximal gastric cancer, metastatic deposits) and less commonly due to extrinsic compression (lymphoma, lung cancer) or paraneoplastic syndromes. Even when not attributable to malignancy, dysphagia is usually a manifestation of an identifiable and treatable disease entity, making its evaluation beneficial to the patient and gratifying to the practitioner. The specific diagnostic algorithm to pursue is guided by the details of the history (Fig. 44-2). If oral or pharyngeal dysphagia is suspected, a fluoroscopic swallow study, usually done by a swallow therapist, is the procedure of choice. Otolaryngoscopic and neurologic evaluation also can be important, depending on the circumstances. For suspected esophageal dysphagia, upper endoscopy is the single most useful test. Endoscopy allows better visualization of mucosal lesions than does barium radiography and also allows for procurement of mucosal biopsies. Endoscopic or histologic abnormalities are evident in the leading causes of esophageal dysphagia: Schatzki's ring, gastroesophageal reflux disease, and eosinophilic esophagitis. Furthermore, therapeutic intervention with esophageal dilation can be done as part of the procedure if it is deemed necessary. The emergence of eosinophilic esophagitis as a leading cause of dysphagia in both children and adults has led to the recommendation that esophageal mucosal biopsies be obtained routinely in the evaluation of unexplained dysphagia even if characteristic, endoscopically identified esophageal mucosal features are absent. For cases of suspected esophageal motility disorders, endoscopy is still the appropriate initial evaluation as neoplastic and inflammatory conditions can secondarily produce patterns of either achalasia or esophageal spasm. Esophageal manometry is done if dysphagia is not adequately explained by endoscopy or to confirm the diagnosis of a suspected esophageal motor disorder. Barium radiography can provide useful adjunctive information in cases of subtle or complex

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 44: Dysphagia, Ikuo Hirano; Peter J. Kahrilas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TREATMENT

Access Provided by:
SILVERCHAIR

diagnosis of a suspected esophageal motor disorder. Barium radiography can provide useful adjunctive information in cases of subtle or complex esophageal strictures, prior esophageal surgery, esophageal diverticula, or paraesophageal herniation. Use of a barium tablet in conjunction with fluoroscopy can identify strictures and esophageal motility disorders that may be overlooked with liquid barium. In specific cases, computed tomography (CT) examination, esophageal manometry with solid meal challenge, and endoscopic ultrasonography may be useful.

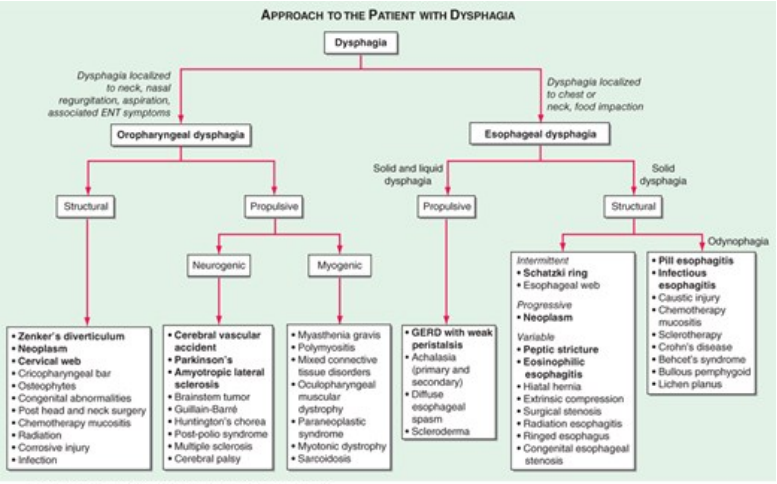
TREATMENT

Treatment of dysphagia depends on both the locus and the specific etiology. Oropharyngeal dysphagia most commonly results from functional deficits caused by neurologic disorders. In such circumstances, the treatment focuses on utilizing postures or maneuvers devised to reduce pharyngeal residue and enhance airway protection learned under the direction of a swallow therapist. Aspiration risk may be reduced by altering the consistency of ingested food and liquid. Dysphagia resulting from a cerebrovascular accident usually, but not always, spontaneously improves within the first few weeks after the event. More severe and persistent cases may require consideration of gastrostomy and enteral feeding. Patients with myasthenia gravis **(Chap. 448)** and polymyositis **(Chap. 365)** may respond to medical treatment of the primary neuromuscular disease. Surgical intervention with cricopharyngeal myotomy is usually not helpful, with the exception of specific disorders such as symptomatic cricopharyngeal bar, Zenker's diverticulum, and oculopharyngeal muscular dystrophy. Chronic neurologic disorders such as Parkinson's disease and amyotrophic lateral sclerosis may manifest with severe oropharyngeal dysphagia. Feeding by a nasogastric tube or an endoscopically placed gastrostomy tube may be considered for nutritional support; however, these maneuvers do not provide protection against aspiration of salivary secretions or refluxed gastric contents.

Treatment of esophageal dysphagia is covered in detail in **Chap. 323**. The majority of causes of structural, esophageal dysphagia are effectively managed by means of esophageal dilation using bougie or balloon dilators. Cancer and achalasia are often managed surgically, although endoscopic techniques are available for both palliation and primary therapy, respectively. Infectious etiologies respond to antimicrobial medications or treatment of the underlying immunosuppressive state. Finally, eosinophilic esophagitis is an important and increasingly recognized cause of dysphagia that is amenable to treatment by elimination of dietary allergens, proton pump inhibition or swallowed, topically acting glucocorticoids in combination with esophageal dilation for persistent strictures.

FIGURE 44-2

Approach to the patient with dysphagia. Etiologies in bold print are the most common. ENT, ear, nose, and throat; GERD, gastroesophageal reflux disease.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FURTHER READING

Hirano I: Esophagus: Anatomy and structural anomalies, in *Yamada Atlas of Gastroenterology*, 6th ed. New York, Wiley-Blackwell Publishing Co., 2016, pp 42–59.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 44: Dysphagia, Ikuo Hirano; Peter J. Kahrilas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Pandolfino JP, Kahrilas PJ: Esophageal neuromuscular function and motility disorders, in *Sleisenger and Fordtran's Gastrointestinal and Liver*

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

pp 42–55.

Kahrilas PJ et al: The Chicago Classification of esophageal motility disorders, v3.0. Neurogastroenterol Motil 27:160, 2015. [PubMed: 25469569]

Kim JP, Kahrilas PJ: How I approach dysphagia. Curr Gastroenterol Rep 21:49, 2019. [PubMed: 31432250]

Pandolfino JP, Kahrilas PJ: Esophageal neuromuscular function and motility disorders, in *Sleisenger and Fordtran's Gastrointestinal and Liver Disease*, 10th ed, Feldman M, Friedman LS, Brandt LJ (eds). Philadelphia, Elsevier, 2016, pp 701–732.

Shaker R et al (eds): *Principles of Deglutition: A Multidisciplinary Text for Swallowing and Its Disorders*. New York, Springer, 2013.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 44: Dysphagia, Ikuo Hirano; Peter J. Kahrilas
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 45: Nausea, Vomiting, and Indigestion

William L. Hasler

INTRODUCTION

Nausea is the feeling of a need to vomit. *Vomiting* (emesis) is the oral expulsion of gastrointestinal contents resulting from gut and thoracoabdominal wall contractions. Vomiting is contrasted with *regurgitation,* the effortless passage of gastric contents into the mouth. *Rumination* is the repeated regurgitation of food residue, which may be rechewed and reswallowed. In contrast to emesis, these phenomena exhibit volitional control. *Indigestion* encompasses a range of complaints including nausea, vomiting, heartburn, regurgitation, and dyspepsia (symptoms thought to originate in the gastroduodenal region). Some individuals with dyspepsia experience postprandial fullness, early satiety (inability to complete a meal due to premature fullness), bloating, eructation (belching), and anorexia. Others report predominantly epigastric burning or pain. Nausea, vomiting, and dyspepsia have been correlated with a condition now called avoidant/restrictive food intake disorder.

NAUSEA AND VOMITING

MECHANISMS

Vomiting is coordinated by the brainstem and is effected by responses in the gut, pharynx, and somatic musculature. Mechanisms underlying nausea are poorly understood but likely involve the cerebral cortex, as nausea requires cognitive and emotional input and is associated with autonomic responses including diaphoresis, pallor, and altered heart rate. Functional brain imaging studies support this idea showing activation of cerebral regions including the insula, anterior cingulate cortex, and amygdala during nausea.

Coordination of Emesis

Brainstem nuclei—including the nucleus tractus solitarius; dorsal vagal and phrenic nuclei; medullary nuclei regulating respiration; and nuclei that control pharyngeal, facial, and tongue movements—coordinate initiation of emesis involving neurokinin NK_1, serotonin 5-HT_3, endocannabinoid, and vasopressin pathways.

Somatic and visceral muscles respond stereotypically during emesis. Inspiratory thoracic and abdominal wall muscles contract, increasing intrathoracic and intraabdominal pressures to evacuate the stomach. Under normal conditions, distally migrating gut contractions are coordinated by an electrical phenomenon, the slow wave, which cycles at 3 cycles/min in the stomach and 11 cycles/min in the duodenum. During emesis, slow waves are abolished and replaced by orally propagating spikes that evoke retrograde contractions to facilitate expulsion of gut contents.

Activators of Emesis

Emetic stimuli act at several sites. Emesis evoked by unpleasant thoughts or smells originates in the brain. Motion sickness and inner ear disorders act on labyrinthine pathways. Gastric irritants and cytotoxic agents like cisplatin stimulate gastroduodenal vagal afferent nerves. Nongastric afferents are activated by bowel obstruction and mesenteric ischemia. The area postrema, in the medulla, responds to bloodborne stimuli (emetogenic drugs, bacterial toxins, uremia, hypoxia, ketoacidosis) and is termed the *chemoreceptor trigger zone*.

Neurotransmitters mediating vomiting are selective for different sites. Labyrinthine disorders stimulate vestibular muscarinic M_1 and histaminergic H_1 receptors. Vagal afferent stimuli activate 5-HT_3 receptors. The area postrema is served by nerves acting on 5-HT_3, M_1, H_1, and dopamine D_2 subtypes. NK_1 receptors in the central nervous system (CNS) mediate both nausea and vomiting. Cannabinoid CB_1 pathways may participate in the cerebral cortex and brainstem. Therapies for vomiting act on these receptor-mediated pathways.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

vomiting is reported by 2–3% of the population.

DIFFERENTIAL DIAGNOSIS

Nausea and vomiting are caused by conditions within and outside the gut, drugs, and circulating toxins **(Table 45-1)**. Unexplained chronic nausea and vomiting is reported by 2–3% of the population.

TABLE 45-1

Causes of Nausea and Vomiting

INTRAPERITONEAL	EXTRAPERITONEAL	MEDICATIONS/METABOLIC DISORDERS
Obstructing disorders	Cardiopulmonary disease	Drugs
Pyloric obstruction	Cardiomyopathy	Cancer chemotherapy
Small-bowel obstruction	Myocardial infarction	Analgesics
Colonic obstruction	Labyrinthine disease	Opioids
Superior mesenteric artery syndrome	Motion sickness	Antibiotics
Enteric infections	Labyrinthitis	Cardiac antiarrhythmics
Viral	Malignancy	Digoxin
Bacterial	Intracerebral disorders	Oral hypoglycemics
Inflammatory diseases	Malignancy	Oral contraceptives
Cholecystitis	Hemorrhage	Antidepressants
Pancreatitis	Abscess	Restless legs/Parkinson's therapies
Appendicitis	Hydrocephalus	Smoking cessation agents
Hepatitis	Psychiatric illness	Endocrine/metabolic disease
Altered sensorimotor function	Anorexia and bulimia nervosa	Pregnancy
Gastroparesis	Depression	Uremia
Intestinal pseudoobstruction	Postoperative vomiting	Ketoacidosis
Gastroesophageal reflux		Thyroid and parathyroid disease
Chronic nausea vomiting syndrome		Adrenal insufficiency
Cyclic vomiting syndrome		Toxins
Cannabinoid hyperemesis syndrome		Liver failure
Rumination syndrome		Ethanol
Mesenteric insufficiency		
Celiac artery stenosis		
Median arcuate ligament syndrome		
Biliary colic		
Abdominal irradiation		

Intraperitoneal Disorders

Obstruction and inflammation of hollow and solid viscera may elicit vomiting. Ulcers and malignancy cause gastric obstruction, while adhesions, benign or malignant tumors, volvulus, intussusception, or inflammatory diseases like Crohn's disease cause small intestinal and colonic obstruction. The superior mesenteric artery syndrome, occurring after weight loss or prolonged bed rest, results when the duodenum is compressed by the overlying superior mesenteric artery. Median arcuate ligament syndrome, with compression of the celiac artery, is a rare cause of vomiting. Abdominal irradiation impairs intestinal motility and induces strictures. Biliary colic causes nausea by acting on afferent nerves. Vomiting with pancreatitis, cholecystitis, and appendicitis results from visceral irritation and induction of ileus. Enteric infectious causes of vomiting include viruses (norovirus, rotavirus), bacteria (*Staphylococcus aureus*, *Bacillus cereus*), and opportunistic organisms like cytomegalovirus or herpes simplex in immunocompromised individuals.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

20% of cases. Rapid gastric emptying is associated with nausea and vomiting in some conditions. *Intestinal pseudoobstruction* is characterized by disrupted intestinal motility with retention of food residue and secretions; bacterial overgrowth; nutrient malabsorption; and symptoms of nausea,

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

immunocompromised individuals.

Gut sensorimotor dysfunction often causes nausea and vomiting. *Gastroparesis* presents with these symptoms with evidence of delayed gastric emptying and occurs after vagotomy or with pancreatic carcinoma, mesenteric vascular insufficiency, or organic diseases like diabetes, scleroderma, and amyloidosis. Idiopathic gastroparesis is the most prevalent etiology; it occurs in the absence of systemic illness and follow a viral illness in ~15–20% of cases. Rapid gastric emptying is associated with nausea and vomiting in some conditions. *Intestinal pseudoobstruction* is characterized by disrupted intestinal motility with retention of food residue and secretions; bacterial overgrowth; nutrient malabsorption; and symptoms of nausea, vomiting, bloating, pain, and altered defecation. Intestinal pseudoobstruction may be idiopathic, inherited, result from systemic disease like scleroderma or an infiltrative process like amyloidosis, or occur as a paraneoplastic consequence of malignancy (e.g., small-cell lung carcinoma). Patients with gastroesophageal reflux, irritable bowel syndrome (IBS), or chronic constipation often report nausea and vomiting.

Other functional gastroduodenal disorders without organic abnormalities have been characterized. *Chronic nausea vomiting syndrome* is defined as bothersome nausea at least 1 day and/or one or more vomiting episodes weekly in the absence of an eating disorder or psychiatric disease. *Cyclic vomiting syndrome (CVS)* causes 3–14% of cases of unexplained nausea and vomiting and presents with discrete episodes of relentless vomiting and is associated with migraines. Some adult cases have been associated with rapid gastric emptying. A related condition, *cannabinoid hyperemesis syndrome (CHS)*, presents with cyclical vomiting in individuals (mostly men) with long-standing use of large quantities of cannabis and resolves with its discontinuation. *Rumination syndrome* is often misdiagnosed as refractory vomiting.

Extraperitoneal Disorders

Myocardial infarction and congestive heart failure may cause nausea and vomiting. Postoperative emesis occurs after 25% of surgeries, especially abdominal and orthopedic surgery. Increased intracranial pressure from tumors, bleeding, abscess, or blockage of cerebrospinal fluid outflow produces vomiting with or without nausea. Patients with anorexia nervosa, bulimia nervosa, anxiety, and depression often report significant nausea associated with delayed gastric emptying.

Medications and Metabolic Disorders

Drugs evoke vomiting by action on the stomach (analgesics, erythromycin) or area postrema (opioids, anti-parkinsonian drugs). Other emetogenic agents include antibiotics, cardiac antiarrhythmics, antihypertensives, oral hypoglycemics, antidepressants (selective serotonin and serotonin norepinephrine reuptake inhibitors), smoking cessation drugs (varenicline, nicotine), and contraceptives. Cancer chemotherapy causes acute (within hours of administration), delayed (after 1 or more days), or anticipatory vomiting. Acute emesis from highly emetogenic agents (e.g., cisplatin) is mediated by 5-HT_3 pathways. Delayed emesis is more dependent on NK_1 mechanisms. Anticipatory nausea may respond to anxiolytic therapy rather than antiemetics.

Metabolic disorders elicit nausea and vomiting. Nausea affects 70% of women in the first trimester of pregnancy. Hyperemesis gravidarum is a severe form of nausea of pregnancy that produces dehydration and electrolyte disturbances and has been proposed to result from excessive amounts of a blood protein—growth differentiation factor 15. Uremia, ketoacidosis, adrenal insufficiency, and parathyroid and thyroid disease are other metabolic etiologies.

Circulating toxins evoke emesis via effects on the area postrema. Endogenous toxins are generated in fulminant liver failure, whereas exogenous enterotoxins may be produced by enteric bacterial infection. Ethanol intoxication is a common toxic etiology of nausea and vomiting.

APPROACH TO THE PATIENT

Nausea and Vomiting

HISTORY AND PHYSICAL EXAMINATION

The history helps define the etiology of nausea and vomiting. Drugs, toxins, and infections often cause acute symptoms, whereas established illnesses evoke chronic complaints. Gastroparesis and pyloric obstruction elicit vomiting within an hour of eating. Emesis from intestinal blockage occurs later. Vomiting occurring minutes after meal consumption prompts consideration of rumination syndrome. With severe gastric emptying delays, vomitus may contain food residue ingested days before. Hematemesis raises suspicion of ulcer, malignancy, or Mallory-Weiss tear. Feculent emesis is noted with distal intestinal or colonic obstruction. Bilious vomiting excludes gastric obstruction, whereas emesis of undigested food is consistent with a Zenker's diverticulum or achalasia. Vomiting can relieve abdominal pain from a bowel obstruction but has no effect in pancreatitis

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The physical examination complements the history. Orthostatic hypotension and reduced skin turgor indicate intravascular fluid loss. Pulmonary abnormalities raise concern for aspiration of vomitus. Bowel sounds are absent with ileus. High-pitched rushes suggest bowel obstruction, whereas a succussion splash is found with gastroparesis or pyloric obstruction. Involuntary guarding raises suspicion of inflammation. Focal blood suggests

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

APPROACH TO THE PATIENT

Nausea and Vomiting

HISTORY AND PHYSICAL EXAMINATION

The history helps define the etiology of nausea and vomiting. Drugs, toxins, and infections often cause acute symptoms, whereas established illnesses evoke chronic complaints. Gastroparesis and pyloric obstruction elicit vomiting within an hour of eating. Emesis from intestinal blockage occurs later. Vomiting occurring minutes after meal consumption prompts consideration of rumination syndrome. With severe gastric emptying delays, vomitus may contain food residue ingested days before. Hematemesis raises suspicion of ulcer, malignancy, or Mallory-Weiss tear. Feculent emesis is noted with distal intestinal or colonic obstruction. Bilious vomiting excludes gastric obstruction, whereas emesis of undigested food is consistent with a Zenker's diverticulum or achalasia. Vomiting can relieve abdominal pain from a bowel obstruction but has no effect in pancreatitis or cholecystitis. Weight loss raises concern about malignancy. Taking prolonged hot baths or showers is associated with CHS and CVS. Intracranial sources are considered if there are headaches or visual changes. Vertigo or tinnitus indicates labyrinthine disease.

The physical examination complements the history. Orthostatic hypotension and reduced skin turgor indicate intravascular fluid loss. Pulmonary abnormalities raise concern for aspiration of vomitus. Bowel sounds are absent with ileus. High-pitched rushes suggest bowel obstruction, whereas a succussion splash is found with gastroparesis or pyloric obstruction. Involuntary guarding raises suspicion of inflammation. Fecal blood suggests ulcer, ischemia, or tumor. Neurologic disease presents with papilledema, visual loss, or focal neural abnormalities. Neoplasm is suggested by palpable masses or adenopathy.

DIAGNOSTIC TESTING

For intractable symptoms or an elusive diagnosis, screening testing can direct care. Electrolyte replacement is indicated for hypokalemia or metabolic alkalosis. Iron-deficiency anemia mandates exclusion of mucosal causes. Abnormal pancreatic or liver biochemistries are found with pancreaticobiliary disease. Endocrinologic, rheumatologic, or paraneoplastic etiologies are suggested by hormone or serologic abnormalities. Supine and upright abdominal radiographs may show intestinal air-fluid levels and reduced colonic air with small-bowel obstruction. Ileus is characterized by diffusely dilated air-filled bowel loops.

Anatomic studies are indicated if initial testing is nondiagnostic. Upper endoscopy detects ulcers, malignancy, and retained food in gastroparesis. Small-bowel barium radiography or computed tomography (CT) diagnoses partial bowel obstruction. Colonoscopy or contrast enema radiography detects colonic obstruction. Ultrasound or CT defines intraperitoneal inflammation; CT and magnetic resonance imaging (MRI) enterography define inflammation in Crohn's disease. Brain CT or MRI delineates intracranial disease. Mesenteric angiography, CT, or MRI is useful for suspected ischemia.

Gastrointestinal motility testing can detect an underlying motor disorder. Gastroparesis commonly is diagnosed by gastric scintigraphy, which measures emptying of a radiolabeled meal. A nonradioactive ^{13}C-labeled gastric emptying breath test is an alternative to scintigraphy. Intestinal pseudoobstruction is suggested by luminal dilation on imaging or abnormal transit on contrast radiography or intestinal scintigraphy. Wireless motility capsules diagnose gastroparesis or small-bowel dysmotility by detecting local or generalized transit delays in the stomach or small bowel from characteristic pH changes between regions. Small-intestinal manometry confirms a diagnosis of pseudoobstruction and discriminates between neuropathic or myopathic disease based on contractile patterns. Manometry can obviate the need for surgical intestinal biopsy to detect smooth muscle or neuronal degeneration. Combined ambulatory esophageal pH/impedance testing and high-resolution manometry facilitates diagnosis of rumination syndrome. Impedance planimetry detects reduced pyloric distensibility in some cases of gastroparesis.

TREATMENT

Nausea and Vomiting

GENERAL PRINCIPLES

Therapy of vomiting is tailored to correct remediable abnormalities if possible. Patients with severe dehydration should be hospitalized if oral fluid replenishment is unsustainable. Once oral intake is tolerated, low-fat liquid nutrients are restarted because lipids delay gastric emptying. Low-residue, small-particle diets have shown efficacy in gastroparesis. Glycemic control should be optimized to reduce diabetic gastroparesis symptoms.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

on vestibular pathways to treat motion sickness and labyrinthine disorders. D_2 antagonists treat emesis evoked by area postrema stimuli including medications, toxins, and metabolic disturbances. Dopamine antagonists cross the blood-brain barrier and cause anxiety, movement disorders, and

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

replenishment is unsustainable. Once oral intake is tolerated, low-fat liquid nutrients are restarted because lipids delay gastric emptying. Low-residue, small-particle diets have shown efficacy in gastroparesis. Glycemic control should be optimized to reduce diabetic gastroparesis symptoms.

ANTIEMETIC MEDICATIONS

Most antiemetic agents act on CNS sites **(Table 45-2)**. Antihistamines like dimenhydrinate and meclizine and anticholinergics like scopolamine act on vestibular pathways to treat motion sickness and labyrinthine disorders. D_2 antagonists treat emesis evoked by area postrema stimuli including medications, toxins, and metabolic disturbances. Dopamine antagonists cross the blood-brain barrier and cause anxiety, movement disorders, and hyperprolactinemic effects (galactorrhea, sexual dysfunction).

Other classes exhibit antiemetic properties. 5-HT_3 antagonists like ondansetron and granisetron prevent postoperative vomiting, radiation therapy–induced symptoms, and cancer chemotherapy–induced emesis, but also are used for other conditions. NK_1 antagonists like aprepitant are approved for chemotherapy-induced vomiting. Aprepitant reduces gastroparesis symptoms. Tricyclic antidepressants reduce symptoms in some patients with functional causes of vomiting, but did not show benefits in a controlled trial in gastroparesis. Other antidepressants such as mirtazapine and olanzapine and the pain-modulating agent gabapentin also exhibit antiemetic effects in some clinical settings.

GASTROINTESTINAL MOTOR STIMULANTS

Drugs that stimulate gastric emptying are used for gastroparesis (Table 45-2). Metoclopramide, a combined 5-HT_4 agonist and D_2 antagonist, is effective in gastroparesis, but antidopaminergic side effects, including dystonias and mood disturbances, limit use in ~25% of cases. Erythromycin increases gastroduodenal motility by action on receptors for motilin, an endogenous transmitter that regulates fasting motility. Intravenous erythromycin is useful for inpatients with refractory gastroparesis. Benefits of long-term oral erythromycin are limited by development of tolerance. Domperidone, a D_2 antagonist not available in the United States, exhibits prokinetic and antiemetic effects but does not cross into most brain regions. The drug rarely causes dystonic reactions but can induce hyperprolactinemic side effects via penetration of pituitary regions served by a porous blood-brain barrier. Prucalopride, a 5-HT_4 agonist, has shown efficacy in accelerating gastric emptying and improving symptoms in idiopathic gastroparesis.

Refractory motility disorders pose challenges. Intestinal pseudoobstruction may respond to the somatostatin analogue octreotide, which induces propagative small-intestinal motor complexes. Acetylcholinesterase inhibitors like pyridostigmine benefit some patients with small-bowel dysmotility. Pyloric botulinum toxin injections reduced gastroparesis symptoms in uncontrolled studies, but small controlled trials observed benefits no greater than sham treatments. Surgical pyloroplasty and gastric peroral endoscopic myotomy (G-POEM) of the pylorus improved symptoms in case series. Enteral feedings through a jejunostomy reduce hospitalizations and improve overall health in some patients with refractory gastroparesis. Subtotal gastric resection may improve some cases of postvagotomy gastroparesis, but its utility for other gastroparesis etiologies is unproven. Implanted gastric electrical stimulators may reduce symptoms, enhance nutrition, improve quality of life, and decrease health care expenditures in medication-refractory gastroparesis; a controlled trial has confirmed modest improvement in vomiting.

SAFETY CONSIDERATIONS

Safety concerns have been raised about selected antiemetics. Metoclopramide can cause irreversible movement disorders like tardive dyskinesia, particularly in older patients. This complication should be explained and documented in the medical record. Domperidone, erythromycin, tricyclic antidepressants, and 5-HT_3 antagonists increase risk of cardiac arrhythmias and sudden cardiac death in those with QTc interval prolongation on electrocardiography (ECG). Surveillance ECG testing is advocated for some of these agents.

OTHER CLINICAL SETTINGS

Some cancer chemotherapies are intensely emetogenic **(Chap. 73)**. Combining a 5-HT_3 antagonist, an NK_1 antagonist, and a glucocorticoid can control both acute and delayed vomiting after highly emetogenic chemotherapy. Benzodiazepines like lorazepam reduce anticipatory nausea and vomiting. Other therapies with benefit in chemotherapy-induced emesis include cannabinoids, olanzapine, gabapentin, and alternative therapies like ginger. Most antiemetic regimens produce greater reductions in chemotherapy-induced vomiting than nausea.

Clinicians should exercise caution in managing nausea of pregnancy. Studies of the teratogenic effects of antiemetic agents provide conflicting results. Antihistamines like meclizine and doxylamine, antidopaminergics like prochlorperazine, and antiserotonergics like ondansetron demonstrate limited efficacy. Some obstetricians recommend alternative therapies including pyridoxine, acupressure, or ginger.

Managing CVS and CHS is challenging. Prophylaxis with tricyclic antidepressants or anticonvulsants (topiramate, zonisamide, levetiracetam) reduces

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

topical capsaicin cream may reduce acute CHS attacks.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Managing CVS and CHS is challenging. Prophylaxis with tricyclic antidepressants or anticonvulsants (topiramate, zonisamide, levetiracetam) reduces the severity and frequency of CVS attacks in uncontrolled reports. Combining intravenous 5-HT_3 antagonists with the sedating effects of a benzodiazepine like lorazepam are mainstays for aborting acute flares. Small studies report benefits with aprepitant and injectable or intranasal forms of the 5-HT_1 agonist sumatriptan to manage acute CVS episodes. These treatments are reportedly less effective for CHS, but haloperidol and topical capsaicin cream may reduce acute CHS attacks.

TABLE 45-2

Treatment of Nausea and Vomiting

TREATMENT	MECHANISM	EXAMPLES	CLINICAL INDICATIONS
Antiemetic agents	Antihistaminergic	Dimenhydrinate, meclizine	Motion sickness, inner ear disease
	Anticholinergic	Scopolamine	Motion sickness, inner ear disease
	Antidopaminergic	Prochlorperazine, thiethylperazine, haloperidol	Medication-, toxin-, or metabolic-induced emesis, chemotherapy-induced nausea and vomiting, ?cannabinoid hyperemesis syndrome
	5-HT_3 antagonist	Ondansetron, granisetron	Chemotherapy- and radiation-induced emesis, postoperative emesis, opioid-induced nausea and vomiting
	Cannabinoids	Tetrahydrocannabinol	Chemotherapy-induced emesis
	Tricyclic antidepressant	Amitriptyline, nortriptyline	Functional vomiting, chronic idiopathic nausea, cyclic vomiting syndrome, ? gastroparesis
	Other antidepressant	Mirtazapine, olanzapine	Functional dyspepsia, ?gastroparesis
	Neuropathic modulator	Gabapentin	Chemotherapy-induced nausea and vomiting
	Neurokinin (NK1) receptor antagonists	Aprepitant, fosaprepitant, netupitant, rolapitant	Chemotherapy-induced emesis
Prokinetic agents	5-HT_4 agonist and antidopaminergic	Metoclopramide	Gastroparesis
	Motilin agonist	Erythromycin	Gastroparesis, ?intestinal pseudoobstruction
	Peripheral antidopaminergic	Domperidone	Gastroparesis
	Pure 5-HT_4 agonist	Prucalopride	?Idiopathic gastroparesis
	Somatostatin analogue	Octreotide	Intestinal pseudoobstruction
Special settings	Benzodiazepines	Lorazepam	Anticipatory nausea and vomiting with chemotherapy, cyclic vomiting syndrome

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

	Acetylcholinesterase inhibitor	Pyridostigmine	?Small-intestinal dysmotility/pseudoobstruction
Special settings	Benzodiazepines	Lorazepam	Anticipatory nausea and vomiting with chemotherapy, cyclic vomiting syndrome
	5-HT_{1A} agonist	Buspirone	Functional dyspepsia
	Glucocorticoids	Methylprednisolone, dexamethasone	Chemotherapy-induced emesis
	Anticonvulsants	Topiramate, zonisamide, levetiracetam	Cyclic vomiting syndrome
	Antimigraine agents	Sumatriptan	Cyclic vomiting syndrome
	Topical analgesic	Capsaicin cream	?Cannabinoid hyperemesis syndrome
	Atypical antipsychotic agent	Olanzapine	Chemotherapy-induced and breakthrough emesis

Note: ?, indication is uncertain.

INDIGESTION

MECHANISMS

Several mechanisms may contribute to indigestion, including acid reflux, altered gut motility or sensation, inflammation, and microbial processes.

Gastroesophageal Reflux

Gastroesophageal reflux results from many defects. Reduced lower esophageal sphincter (LES) tone causes reflux in scleroderma and pregnancy and may be a factor in some patients without systemic illness. Other cases exhibit frequent transient LES relaxations (TLESRs). Reductions in esophageal body motility or saliva production prolong esophageal fluid clearance. Increased intragastric pressure promotes gastroesophageal reflux with obesity. Many reflux patients have hiatal hernias, and large hernias can increase symptomatic reflux.

Gastric Motor Dysfunction

Disturbed gastric motility may contribute to gastroesophageal reflux in up to one-third of cases. Delayed gastric emptying is found in ~30% of functional dyspeptics, while rapid gastric emptying affects 5%. Impaired gastric fundus relaxation after eating (i.e., accommodation) may underlie selected dyspeptic symptoms like bloating, nausea, and early satiety in ~40% of patients and may predispose to TLESRs and acid reflux.

Visceral Afferent Hypersensitivity

Disturbed gastric sensation is another pathogenic factor in functional dyspepsia. Approximately 35% of dyspeptic patients note discomfort with fundic distention to lower pressures than in healthy controls. Other individuals with dyspepsia exhibit hypersensitivity to chemical stimulation of the stomach with capsaicin or with duodenal acid or lipid perfusion. Some cases of functional heartburn without increased acid or nonacid reflux exhibit heightened perception of normal esophageal acidity.

Immune Activation

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

and may selectively elicit early satiety and epigastric pain. Proliferations in duodenal bacteria were shown to correlate with meal-induced symptoms in functional dyspepsia, suggesting a role for microbiome alterations. Intestinal bile salt release also is proposed to worsen dyspeptic symptoms after

Increases in duodenal epithelial permeability in functional dyspepsia may relate to increases in eosinophils and mast cells adjacent to submucosal neurons. Increased activation of these cells is proposed to contribute to gastric emptying delays and altered sensory function in functional dyspepsia and may selectively elicit early satiety and epigastric pain. Proliferations in duodenal bacteria were shown to correlate with meal-induced symptoms in functional dyspepsia, suggesting a role for microbiome alterations. Intestinal bile salt release also is proposed to worsen dyspeptic symptoms after eating. Both dysbiosis and bile may contribute to mucosal permeability defects.

Other Factors

Helicobacter pylori has a proven etiologic role in peptic ulcer disease but is a minor factor in the genesis of functional dyspepsia. Anxiety and depression may play contributing roles in some functional dyspepsia cases. Functional MRI studies show increased activation of several brain regions, emphasizing CNS contributions. Up to 20% of functional dyspepsia patients report symptom onset after a viral illness, suggesting an infectious trigger. Analgesics cause dyspepsia, whereas nitrates, calcium channel blockers, theophylline, and progesterone promote gastroesophageal reflux. Ethanol, tobacco, and caffeine induce LES relaxation and reflux. Genetic factors predispose to development of reflux and dyspepsia in some cases.

DIFFERENTIAL DIAGNOSIS

Gastroesophageal Reflux Disease

Heartburn or regurgitation is reported weekly by 18–28% of the population, highlighting the prevalence of gastroesophageal reflux disease (GERD). Most cases of heartburn result from excess acid reflux, but reflux of weakly acidic or nonacidic fluid can produce similar symptoms. Alkaline reflux esophagitis elicits GERD symptoms in patients who have had surgery for peptic ulcer disease. Ten percent of patients with heartburn exhibit no acidic or nonacidic esophageal reflux and are considered to have functional heartburn.

Functional Dyspepsia

Approximately 20% of the populace has dyspepsia at least six times yearly, but only 10–20% present to clinicians. Functional dyspepsia, the cause of symptoms in 70–80% of dyspeptic patients, is defined as bothersome postprandial fullness, early satiety, or epigastric pain or burning with symptom onset ≥6 months before diagnosis in the absence of organic cause. Functional dyspepsia is subdivided into postprandial distress syndrome (61% of cases), characterized by meal-induced fullness and early satiety, and epigastric pain syndrome (18% of cases), with epigastric pain or burning that may or may not be meal related. Twenty-one percent of individuals present with overlapping postprandial distress and epigastric pain syndromes. Functional dyspepsia is associated with other functional gut disorders including irritable bowel syndrome and nongastrointestinal disorders like fibromyalgia, chronic fatigue, and anxiety. Most cases follow a benign course, but some with *H. pylori* infection or on nonsteroidal anti-inflammatory drugs (NSAIDs) develop ulcers.

Ulcer Disease

Most GERD patients do not exhibit esophageal injury, but 5% develop esophageal ulcers. Symptoms cannot distinguish nonerosive from erosive or ulcerative esophagitis. A minority of cases of dyspepsia stem from gastric or duodenal ulcers. The most common causes of ulcers are *H. pylori* infection and NSAID use. Other rare causes of gastroduodenal ulcers include Crohn's disease **(Chap. 326)** and Zollinger-Ellison syndrome **(Chap. 324)**, resulting from gastrin overproduction by an endocrine tumor.

Malignancy

Dyspeptic patients may seek care because of fear of cancer, but few cases result from malignancy. Esophageal squamous cell carcinoma occurs most often with long-standing tobacco or ethanol intake. Other risks include prior caustic ingestion, achalasia, and the hereditary disorder tylosis. Esophageal adenocarcinoma usually complicates prolonged acid reflux. Eight to 20% of GERD patients exhibit esophageal intestinal metaplasia, termed *Barrett's metaplasia,* which predisposes to esophageal adenocarcinoma **(Chap. 80)**. Gastric malignancies include adenocarcinoma, which is prevalent in certain Asian societies, and lymphoma.

Other Causes

Opportunistic fungal or viral esophageal infections may produce heartburn but more often cause odynophagia. Other causes of esophageal

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

gastroparesis patients note a predominance of pain rather than nausea and vomiting. Intestinal lactase deficiency may cause gas, bloating, and discomfort and occurs more commonly in blacks and Asians. Intolerance of other carbohydrates (e.g., fructose, sorbitol) produces similar symptoms.

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Opportunistic fungal or viral esophageal infections may produce heartburn but more often cause odynophagia. Other causes of esophageal inflammation include eosinophilic esophagitis and pill esophagitis. Biliary colic is a potential cause unexplained upper abdominal pain, but most patients report discrete acute episodes of right upper quadrant or epigastric pain rather than chronic burning or fullness. Twenty percent of gastroparesis patients note a predominance of pain rather than nausea and vomiting. Intestinal lactase deficiency may cause gas, bloating, and discomfort and occurs more commonly in blacks and Asians. Intolerance of other carbohydrates (e.g., fructose, sorbitol) produces similar symptoms. Small-intestinal bacterial overgrowth may cause dyspepsia, as well as bowel dysfunction, distention, and malabsorption. Celiac disease, nonceliac gluten sensitivity, pancreatic disease (chronic pancreatitis, malignancy), hepatocellular carcinoma, Ménétrier's disease, infiltrative diseases (sarcoidosis, mastocytosis, eosinophilic gastroenteritis), mesenteric ischemia, thyroid and parathyroid disease, and abdominal wall strain cause dyspepsia. Extraperitoneal etiologies of indigestion include congestive heart failure and tuberculosis.

APPROACH TO THE PATIENT

Indigestion

HISTORY AND PHYSICAL EXAMINATION

Managing indigestion requires a thorough interview. GERD classically produces heartburn, a substernal warmth that moves toward the neck. Heartburn often is exacerbated by meals and may awaken the patient. Associated symptoms include regurgitation of acid or nonacidic fluid and water brash, the reflex release of salty saliva into the mouth. Atypical symptoms include pharyngitis, asthma, cough, bronchitis, hoarseness, and chest pain that mimics angina. Some patients with acid reflux on esophageal pH testing note abdominal pain instead of heartburn.

Dyspeptic patients report symptoms referable to the upper abdomen that may be meal-related (postprandial distress syndrome) or independent of food ingestion (epigastric pain syndrome). The history in functional dyspepsia may also report symptoms of GERD, IBS, or idiopathic gastroparesis.

The physical exam with GERD and functional dyspepsia usually is normal. In atypical GERD, pharyngeal erythema and wheezing may be noted. Recurrent regurgitation may cause poor dentition. Dyspeptics may exhibit epigastric tenderness or distention.

Discriminating functional from organic causes of indigestion mandates excluding certain historic and exam features. Odynophagia suggests esophageal infection. Dysphagia is concerning for a benign or malignant esophageal blockage. Other alarm features include unexplained weight loss, recurrent vomiting, dysphagia, occult or gross bleeding, nocturnal symptoms, jaundice, palpable mass or adenopathy, and a family history of gastrointestinal neoplasm. Patients with an abdominal wall source of upper abdominal pain may exhibit a positive Carnett's sign of increased tenderness with tensing of abdominal muscles upon lifting the head from the exam table.

DIAGNOSTIC TESTING

Because indigestion is prevalent and most cases result from GERD or functional dyspepsia, it is generally recommended to perform no more than limited and directed diagnostic testing in most individuals.

After excluding alarm factors **(Table 45-3)**, patients with typical GERD do not need further evaluation and are treated empirically. Upper endoscopy is indicated only in cases with atypical symptoms or these alarm factors. For heartburn >5 years in duration, especially in patients >50 years old, endoscopy is advocated to screen for Barrett's metaplasia. Endoscopy is not needed in low-risk patients who respond to acid suppressants. Ambulatory esophageal pH testing using a catheter method or a wireless capsule endoscopically attached to the esophageal wall is considered for drug-refractory symptoms and atypical symptoms like unexplained chest pain. High-resolution esophageal manometry is ordered when surgical treatment of GERD is considered. A low LES pressure predicts failure of drug therapy and provides a rationale to proceed to surgery. Poor esophageal body peristalsis raises concern about postoperative dysphagia and directs the choice of surgical technique. Nonacidic reflux may be detected by combined esophageal impedance-pH testing in medication-unresponsive patients.

Upper endoscopy is recommended as the initial test in patients with unexplained dyspepsia who are >60 years old to exclude malignancy—a finding in only 0.3% of endoscopies performed for uninvestigated dyspepsia. Management of patients <60 years old depends on the local *H. pylori* prevalence. In regions with low prevalence (<10%), a 4-week trial of an acid-suppressing medication such as a proton pump inhibitor (PPI) is recommended. If empiric acid suppression fails, a "test and treat" approach for *H. pylori* status is initiated with urea breath testing or stool antigen measurement. Those who are *H. pylori* positive are given therapy to eradicate infection. For patients in areas with high *H. pylori* prevalence (>10%), an initial "test and treat" approach is advocated, and empiric PPI therapy is reserved for those who are negative for infection or who fail to respond to *H. pylori* treatment. Patients who are treated for *H. pylori* should undergo confirmation of eradication with repeat urea breath testing or fecal

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Further testing is indicated in some settings. For suspected bleeding, a blood count can exclude anemia. Thyroid chemistries or calcium levels

in only 0.3% of endoscopies performed for uninvestigated dyspepsia. Management of patients <60 years old depends on the local *H. pylori*

ACCESS Medicine

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

an initial "test and treat" approach is advocated, and empiric PPI therapy is reserved for those who are negative for infection or who fail to respond to *H. pylori* treatment. Patients who are treated for *H. pylori* should undergo confirmation of eradication with repeat urea breath testing or fecal antigen testing 4–6 weeks after completing therapy. Those under age 60 only warrant upper endoscopy if their symptoms fail to respond to these therapies. Some advocate initial endoscopy for patients <60 years old who report alarm symptoms, but some guidelines have not endorsed this practice unless symptoms persist despite treatment.

Further testing is indicated in some settings. For suspected bleeding, a blood count can exclude anemia. Thyroid chemistries or calcium levels screen for metabolic disease. Specific serologies may suggest celiac disease. Pancreatic and liver chemistries are obtained for suspected pancreaticobiliary causes, which are further investigated with ultrasound, CT, or MRI. Gastric emptying testing is considered to exclude gastroparesis for dyspeptic symptoms resembling postprandial distress when therapy fails. Breath testing after carbohydrate ingestion detects lactase deficiency, intolerance to other carbohydrates, or small-intestinal bacterial overgrowth.

TABLE 45-3

Alarm Symptoms in Gastroesophageal Reflux Disease

Odynophagia or dysphagia
Unexplained weight loss
Recurrent vomiting
Occult or gross gastrointestinal bleeding
Jaundice
Palpable mass or adenopathy
Family history of gastroesophageal malignancy

TREATMENT

Indigestion

LIFESTYLE, DIET, AND NONMEDICATION RECOMMENDATIONS

Patients with mild indigestion can be reassured that a careful evaluation revealed no serious disease and are offered no other intervention. If possible, drugs that cause gastroesophageal reflux or dyspepsia should be stopped. GERD patients should limit ethanol, caffeine, chocolate, and tobacco use and can ingest a low-fat diet, avoid snacks before bedtime, and elevate the head of the bed. Functional dyspepsia patients can be advised to reduce intake of fat, spicy foods, caffeine, and alcohol. Dietary lactose restriction is appropriate for lactase deficiency, while gluten exclusion is indicated for celiac disease. Low FODMAP (fermentable oligosaccharide, disaccharide, monosaccharide, and polyol) diets are effective for gaseous symptoms in IBS. In a systematic review, FODMAP intake correlated with functional dyspepsia symptoms, suggesting potential utility in this disorder as well.

ACID-SUPPRESSING OR -NEUTRALIZING MEDICATIONS

Drugs that reduce or neutralize gastric acid are often prescribed for GERD. Histamine H_2 antagonists like cimetidine, ranitidine, famotidine, and nizatidine are useful in mild to moderate GERD. For severe symptoms or for many cases of erosive or ulcerative esophagitis, PPIs like omeprazole, lansoprazole, rabeprazole, pantoprazole, esomeprazole, or dexlansoprazole are needed. These drugs inhibit gastric H^+, K^+-ATPase and are more potent than H_2 antagonists. Up to one-third of GERD patients do not respond to standard PPI doses; one-third of these patients have nonacidic reflux, whereas 10% have persistent acid-related disease. Heartburn responds better to PPI therapy than regurgitation or atypical GERD symptoms.

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

hypomagnesemia, bone demineralization, interstitial nephritis, and impaired medication absorption (clopidogrel). Many patients started on a PPI can be stepped down to an H_2 antagonist or switched to on-demand use.

reflux, whereas 10% have persistent acid-related disease. Heartburn responds better to PPI therapy than regurgitation or atypical GERD symptoms. Some individuals respond to doubling of the PPI dose or adding an H_2 antagonist. Complications of long-term PPI therapy include diarrhea (*Clostridium difficile* infection, microscopic colitis), small-intestinal bacterial overgrowth, nutrient deficiency (vitamin B_{12}, iron, calcium), hypomagnesemia, bone demineralization, interstitial nephritis, and impaired medication absorption (clopidogrel). Many patients started on a PPI can be stepped down to an H_2 antagonist or switched to on-demand use.

Acid suppressants also are effective for both the postprandial distress and epigastric pain subtypes of functional dyspepsia. A meta-analysis of 18 controlled trials calculated a risk ratio of 0.88, with a 95% confidence interval of 0.82–0.94, favoring PPI therapy over placebo in functional dyspepsia. H_2 antagonists also improve symptoms in functional dyspepsia, but a guideline has advocated PPIs over H_2 antagonists as first-line therapies for functional dyspepsia. In addition to acid suppression, PPIs may have the additional action of reducing duodenal eosinophil counts in dyspepsia.

Antacids are useful for short-term control of mild GERD but have less benefit in severe cases unless given at high doses that cause side effects (diarrhea and constipation with magnesium- and aluminum-containing agents, respectively). Alginic acid combined with antacids forms a floating barrier to reflux in patients with upright symptoms. Sucralfate, a salt of aluminum hydroxide and sucrose octasulfate that buffers acid and binds pepsin and bile salts, shows efficacy in GERD similar to H_2 antagonists.

HELICOBACTER PYLORI ERADICATION

H. pylori eradication is indicated for peptic ulcer and mucosa-associated lymphoid tissue gastric lymphoma. The benefits of eradication therapy in functional dyspepsia are limited but are statistically significant. A systematic review of 25 controlled trials calculated a pooled risk ratio of 1.24, with a 95% confidence interval of 1.12–1.37, favoring *H. pylori* eradication over placebo. Most drug combinations **(Chaps. 163 and 324)** include 7–14 days of a PPI with two or three antibiotics with or without bismuth products. *H. pylori* infection is associated with reduced prevalence of GERD. However, eradication of infection does not worsen GERD symptoms. No consensus recommendations regarding *H. pylori* eradication in GERD patients have been offered.

AGENTS THAT MODIFY GASTROINTESTINAL MOTOR ACTIVITY

The γ-aminobutyric acid B (GABA-B) agonist baclofen reduces esophageal exposure to acid and nonacidic fluids by reducing TLESRs by 40%. This drug can be used in patients with refractory acid or nonacid reflux. Several studies have promoted the efficacy of agents that stimulate gastric emptying in functional dyspepsia with 33% relative risk reductions, but publication bias and small sample sizes raise questions about reported benefits of these agents. Some clinicians suggest that patients with the postprandial distress subtype may respond preferentially to such prokinetic drugs. The newer 5-HT_4 agonist prucalopride was reported to reduce symptoms in patients with idiopathic gastroparesis, but no similar studies have been conducted in functional dyspepsia. The 5-HT_{1A} agonists buspirone and tandospirone may improve some functional dyspepsia symptoms by enhancing meal-induced gastric accommodation. Acotiamide stimulates gastric emptying and augments accommodation by enhancing acetylcholine release via muscarinic receptor antagonism and acetylcholinesterase inhibition. This agent is approved for functional dyspepsia in Japan and India.

ANTIDEPRESSANTS

Some patients with refractory functional heartburn may respond to antidepressants in the tricyclic and selective serotonin reuptake inhibitor (SSRI) classes, although studies are limited. Their mechanism of action may involve blunting of visceral pain processing in the brain. In a controlled trial in functional dyspepsia, the tricyclic drug amitriptyline produced symptom reductions, whereas the SSRI escitalopram had no benefit in a three-way comparison with placebo. In another controlled trial in functional dyspepsia, the antidepressant mirtazapine produced superior symptom reductions versus placebo. However, in a meta-analysis of 13 trials, SSRIs and serotonin-norepinephrine reuptake inhibitors showed no benefits in functional dyspepsia.

OTHER OPTIONS

Antireflux surgery (fundoplication) to enhance the barrier function of the LES may be offered to GERD patients who are young and require lifelong therapy, have typical heartburn, are responsive to PPIs, and show acid reflux on pH monitoring. Surgery also is effective for some cases of nonacidic reflux. Individuals who respond less well to fundoplication include those with atypical symptoms, those who have functional heartburn without reflux on testing, or those who have esophageal body motor disturbances. Dysphagia, gas-bloat syndrome, and gastroparesis are long-term complications of fundoplication; ~60% develop recurrent GERD symptoms over time. Magnetic sphincter augmentation may be appropriate for

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

gas and bloating are bothersome in some patients with indigestion and are difficult to treat. Simethicone, activated charcoal, and alpha galactosidase provide benefits in some cases. One trial suggested possible benefits of the nonabsorbable antibiotic rifaximin in functional dyspepsia, while another reported improvement with the probiotic *Lactobacillus gasseri*. Herbal remedies like STW 5 (Iberogast, a mixture of nine

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

complications of fundoplication; ~60% develop recurrent GERD symptoms over time. Magnetic sphincter augmentation may be appropriate for GERD treatment, while endoscopic radiofrequency therapies can be considered for some patients. Other endoscopic options including transoral incisionless fundoplication, endoscopic stapling, and antireflux mucosectomy are not yet advocated.

Gas and bloating are bothersome in some patients with indigestion and are difficult to treat. Simethicone, activated charcoal, and alpha-galactosidase provide benefits in some cases. One trial suggested possible benefits of the nonabsorbable antibiotic rifaximin in functional dyspepsia, while another reported improvement with the probiotic *Lactobacillus gasseri*. Herbal remedies like STW 5 (Iberogast, a mixture of nine herbal agents) and formulations of caraway oil and menthol are useful in some dyspeptic patients. Psychological treatments (e.g., behavioral therapy, psychotherapy, hypnotherapy) may be offered for refractory functional dyspepsia; a meta-analysis of four trials reported benefits in patients with persistent dyspepsia.

FURTHER READING

Gyawali CP et al: ACG Clinical Guidelines: clinical use of esophageal physiologic testing. Am J Gastroenterol 115:1412, 2020. [PubMed: 32769426]

Maret-Ouda J et al: Gastroesophageal reflux disease: a review. JAMA 324:2536, 2020. [PubMed: 33351048]

Sharaf RN et al: Management of cyclic vomiting syndrome in adults: evidence review. Neurogastroenterol Motil 31(Suppl 2):e13605, 2019. [PubMed: 31241818]

Venkatesan T et al: Role of chronic cannabis use: cyclic vomiting syndrome vs. cannabinoid hyperemesis syndrome. Neurogastroenterol Motil 31(Suppl 2):e13606, 2019. [PubMed: 31241817]

Wauters L et al: Novel concepts in the pathophysiology and treatment of functional dyspepsia. Gut 69:591, 2020. [PubMed: 31784469]

Downloaded 2022-1-3 5:48 P Your IP is 141.216.78.40
Chapter 45: Nausea, Vomiting, and Indigestion, William L. Hasler
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 46: Diarrhea and Constipation

Michael Camilleri; Joseph A. Murray

INTRODUCTION

Diarrhea and constipation are exceedingly common and, together, exact an enormous toll in terms of mortality, morbidity, social inconvenience, loss of work productivity, and consumption of medical resources. Worldwide, >1 billion individuals suffer one or more episodes of acute diarrhea each year. Among the 100 million persons affected annually by acute diarrhea in the United States, nearly half must restrict activities, 10% consult physicians, ~250,000 require hospitalization, and ~5000 die (primarily the elderly). Updated 2014–2015 annual disease burden data from the United States show 3.4 million annual clinic or emergency department visits, about 130,000 hospital admissions, and annual economic burden to society (excluding all costs for inflammatory bowel disease) exceeding $8 billion. Acute infectious diarrhea remains one of the most common causes of mortality in developing countries, particularly among impoverished infants, accounting for 1.8 million deaths per year. Recurrent, acute diarrhea in children in tropical countries results in environmental enteropathy with long-term impacts on physical and intellectual development.

Constipation, by contrast, is rarely associated with mortality and is exceedingly common in developed countries, leading to frequent self-medication and, in a third of those, to medical consultation. Annual disease burden data for 2014–2015 show about 5 million clinic or emergency department visits for constipation or hemorrhoids, 50,000 admissions to hospital, and average cost of $3500 per patient, about double that of controls in a nested controlled study.

Population statistics on chronic diarrhea and constipation are more uncertain, perhaps due to variable definitions and reporting, but the frequency of these conditions is also high. U.S. population surveys put prevalence rates for chronic diarrhea at 2–7% and for chronic constipation at 12–19%, with women being affected twice as often as men, reaching parity at 70 years of age. Diarrhea and constipation are among the most common patient complaints presenting in primary care and account for nearly 50% of referrals to gastroenterologists.

Although diarrhea and constipation may present as mere nuisance symptoms at one extreme, they can be severe or life threatening at the other. Even mild symptoms may signal a serious underlying gastrointestinal (GI) lesion, such as colorectal cancer, or systemic disorder, such as thyroid disease. Given the heterogeneous causes and potential severity of these common complaints, it is imperative for clinicians to appreciate the pathophysiology, etiologic classification, diagnostic strategies, and principles of management of diarrhea and constipation so that rational and cost-effective care can be delivered.

NORMAL PHYSIOLOGY

While the primary function of the small intestine is the digestion and assimilation of nutrients from food, the small intestine and colon together perform important functions that regulate the secretion and absorption of water and electrolytes, the storage and subsequent transport of intraluminal contents aborally, and the salvage of some nutrients that are not absorbed in the small intestine after bacterial metabolism of carbohydrate allows salvage of short-chain fatty acids. The main motor functions are summarized in **Table 46-1**. Alterations in fluid and electrolyte handling contribute significantly to diarrhea. Alterations in motor and sensory functions of the colon result in highly prevalent syndromes such as irritable bowel syndrome (IBS), chronic diarrhea, and chronic constipation.

TABLE 46-1

Normal Gastrointestinal Motility: Functions at Different Anatomic Levels

Stomach and Small Bowel
Synchronized MMC in fasting
Stomach ~3 h

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 46-1

Normal Gastrointestinal Motility: Functions at Different Anatomic Levels

Stomach and Small Bowel
Synchronized MMC in fasting
Accommodation, trituration, mixing, transit
Stomach ~3 h
Small bowel ~3 h
Ileal reservoir empties boluses
Colon: Irregular Mixing, Fermentation, Absorption, Transit
Ascending, transverse: reservoirs
Descending: conduit
Sigmoid/rectum: volitional reservoir

Abbreviation: MMC, migrating motor complex.

NEURAL CONTROL

The small intestine and colon have intrinsic and extrinsic innervation. The *intrinsic innervation*, also called the enteric nervous system, comprises myenteric, submucosal, and mucosal neuronal layers. The function of these layers is modulated by interneurons through the actions of neurotransmitter amines or peptides, including acetylcholine, vasoactive intestinal peptide (VIP), opioids, norepinephrine, serotonin, adenosine triphosphate (ATP), and nitric oxide (NO). The myenteric plexus regulates smooth-muscle function through intermediary pacemaker-like cells called the interstitial cells of Cajal, and the submucosal plexus affects secretion, absorption, and mucosal blood flow. The enteric nervous system receives input from the extrinsic nerves, but it is capable of independent control of these functions.

The *extrinsic innervations* of the small intestine and colon are part of the autonomic nervous system and also modulate motor and secretory functions. The parasympathetic nerves convey visceral sensory pathways from and excitatory pathways to the small intestine and colon. Parasympathetic fibers via the vagus nerve reach the small intestine and proximal colon along the branches of the superior mesenteric artery. The distal colon is supplied by sacral parasympathetic nerves (S_{2-4}) via the pelvic plexus; these fibers course through the wall of the colon as ascending intracolonic fibers as far as, and in some instances including, the proximal colon. The chief excitatory neurotransmitters controlling motor function are acetylcholine and the tachykinins, such as substance P. The sympathetic nerve supply modulates motor functions and reaches the small intestine and colon alongside their arterial vessels. Sympathetic input to the gut is generally excitatory to sphincters and inhibitory to nonsphincteric muscle. Visceral afferents convey sensation from the gut to the central nervous system (CNS). Some afferent fibers synapse in the prevertebral ganglia and reflexly modulate intestinal motility, blood flow, and secretion.

INTESTINAL FLUID ABSORPTION AND SECRETION

On an average day, 9 L of fluid enter the GI tract, ~1 L of residual fluid reaches the colon, and the stool excretion of fluid constitutes about 0.2 L/d. The colon has a large capacitance and functional reserve and may recover up to four times its usual volume of 0.8 L/d, provided the rate of flow permits reabsorption to occur. Thus, the colon can partially compensate for excess fluid delivery to the colon that may result from intestinal absorptive or secretory disorders.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

functions of the basolateral sodium pump. There are several active transport proteins at the apical membrane, especially in the small intestine, whereby sodium ion entry is coupled to monosaccharides (e.g., glucose through the transporter SGLT1, or fructose through GLUT-5). Glucose then

ACCESS Medicine®

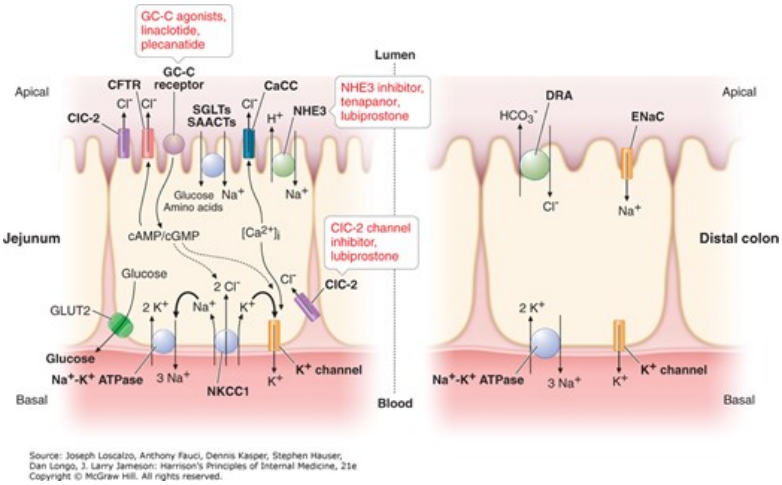
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

secretory disorders.

In the small intestine and colon, sodium absorption is predominantly electrogenic (i.e., it can be measured as an ionic current across the membrane because there is not an equivalent loss of a cation from the cell), and uptake takes place at the apical membrane; it is compensated for by the export functions of the basolateral sodium pump. There are several active transport proteins at the apical membrane, especially in the small intestine, whereby sodium ion entry is coupled to monosaccharides (e.g., glucose through the transporter SGLT1, or fructose through GLUT-5). Glucose then exits the basal membrane through a specific transport protein, GLUT-2, creating a glucose concentration and osmotic gradient between the lumen and the intercellular space, drawing water and electrolytes passively from the lumen. Several channels mediate the secretion of chloride ions in diarrheal diseases or in response to medications administered for the treatment of constipation. The diverse ion channels (chloride channels and cystic fibrosis transmembrane regulator), transporters (SGLT1, GLUT-2), and receptors (e.g., guanylate cyclase C receptor) are summarized in **Figure 46-1.**

FIGURE 46-1

Important ion transport mechanisms in the jejunum and colon, and the site of action of medications used as secretagogues in the treatment of chronic constipation. CFTR, cystic fibrosis transmembrane regulator; ClC2, type 2 chloride channel, DRA, downregulated in adenoma (also called SLC26A3); ENaC, epithelial sodium channel; GC-C, guanylate cyclase C; Na^+-K^+ ATPase, sodium potassium adenosine triphosphatase; NHE3, sodium-hydrogen exchanger; NKCC1, Na-K-Cl cotransporter; SAACT, sodium amino acid co- transporters; SGLT, sodium glucose transporters.



A variety of neural and nonneural mediators regulate colonic fluid and electrolyte balance, including cholinergic, adrenergic, and serotonergic mediators. Angiotensin and aldosterone also influence colonic absorption, reflecting the common embryologic development of the distal colonic epithelium and the renal tubules.

SMALL-INTESTINAL MOTILITY

During the fasting period, the motility of the small intestine is characterized by a cyclical event called the migrating motor complex (MMC), which serves to clear nondigestible residue from the small intestine (the intestinal "housekeeper"). This organized, propagated series of contractions lasts, on average, 4 min, occurs every 60–90 min, and usually involves the entire small intestine. After food ingestion, the small intestine produces irregular, mixing contractions of relatively low amplitude, except in the distal ileum where more powerful contractions occur intermittently and empty the ileum by bolus transfers.

ILEOCOLONIC STORAGE AND SALVAGE

The distal ileum acts as a reservoir, emptying intermittently by bolus movements. This action allows time for salvage of fluids, electrolytes, and nutrients. Segmentation by haustra compartmentalizes the colon and facilitates mixing, retention of residue, and formation of solid stools. There is increased appreciation of the intimate interaction between the colonic function and the luminal ecology. The resident microorganisms, predominantly anaerobic bacteria, in the colon are necessary for the digestion of unabsorbed carbohydrates that reach the colon even in health, thereby providing a vital source of nutrients to the mucosa. Normal intestinal flora also keeps pathogens at bay by a variety of mechanisms including a crucial role in the development and maintenance of a potent but well-regulated immune response capacity to pathogens and tolerance to normal ingesta. In health, the ascending and transverse regions of colon function as reservoirs (average transit time, 15 h), and the descending colon acts as a conduit (average

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

proximal colon or the propulsive function of the left colon. Constipation may also result from disturbances of the rectal or sigmoid reservoir, typically as a result of dysfunction of the pelvic floor, the anal sphincters, the coordination of defecation, or dehydration.

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

development and maintenance of a potent but well-regulated immune response capacity to pathogens and tolerance to normal ingesta. In health, the ascending and transverse regions of colon function as reservoirs (average transit time, 15 h), and the descending colon acts as a conduit (average transit time, 3 h). The colon is efficient at conserving sodium and water, a function that is particularly important in sodium-depleted patients in whom the small intestine alone is unable to maintain sodium balance. Diarrhea or constipation may result from alteration in the reservoir function of the proximal colon or the propulsive function of the left colon. Constipation may also result from disturbances of the rectal or sigmoid reservoir, typically as a result of dysfunction of the pelvic floor, the anal sphincters, the coordination of defecation, or dehydration.

COLONIC MOTILITY AND TONE

The small-intestinal MMC only rarely continues into the colon. However, short duration or phasic contractions mix colonic contents, and high-amplitude (>75 mmHg) propagated contractions (HAPCs) are sometimes associated with mass movements through the colon and normally occur approximately five times per day, usually on awakening in the morning and postprandially. Increased frequency of HAPCs may result in diarrhea or urgency. The predominant phasic contractions in the colon are irregular and nonpropagated and serve a “mixing” function.

Colonic tone refers to the background contractility upon which phasic contractile activity (typically contractions lasting <15 s) is superimposed. It is an important cofactor in the colon’s capacitance (volume accommodation) and sensation.

COLONIC MOTILITY AFTER MEAL INGESTION

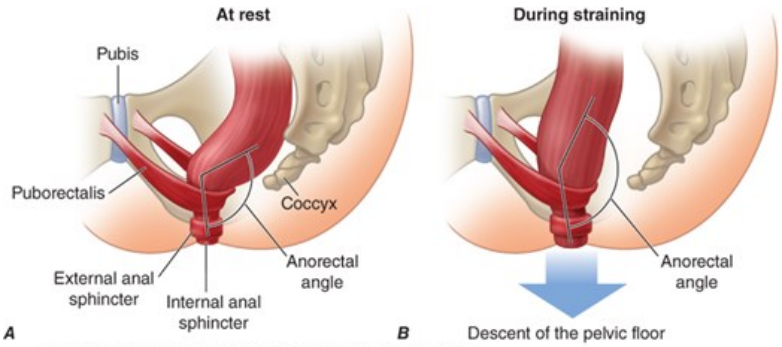
After meal ingestion, colonic phasic and tonic contractility increase for a period of ~2 h. The initial phase (~10 min) is mediated by the vagus nerve in response to mechanical distention of the stomach. The subsequent response of the colon requires caloric stimulation (e.g., intake of at least 500 kcal) and is mediated, at least in part, by hormones (e.g., gastrin and serotonin).

DEFECATION

Tonic contraction of the puborectalis muscle, which forms a sling around the rectoanal junction, is important to maintain continence; during defecation, sacral parasympathetic nerves relax this muscle, facilitating the straightening of the rectoanal angle **(Fig. 46-2)**. Distention of the rectum results in transient relaxation of the internal anal sphincter via intrinsic and reflex sympathetic innervation. As sigmoid and rectal contractions, as well as straining (Valsalva maneuver), which increases intraabdominal pressure, increase the pressure within the rectum, the rectosigmoid angle opens by >15°. Voluntary relaxation of the external anal sphincter (striated muscle innervated by the pudendal nerve) in response to the sensation produced by distention permits the evacuation of feces. Defecation can also be delayed voluntarily by contraction of the external anal sphincter.

FIGURE 46-2

Sagittal view of the anorectum (*A*) at rest and (*B*) during straining to defecate. Continence is maintained by normal rectal sensation and tonic contraction of the internal anal sphincter and the puborectalis muscle, which wraps around the anorectum, maintaining an anorectal angle between 80° and 110°. During defecation, the pelvic floor muscles (including the puborectalis) relax, allowing the anorectal angle to straighten by at least 15°, and the perineum descends by 1–3.5 cm. The external anal sphincter also relaxes and reduces pressure on the anal canal. *(From N Engl J Med, A Lembo, M Camilleri, Chronic Constipation, 349:1360, 2003 Massachusetts Medical Society. Reprinted with permission.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

DIARRHEA

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

weight >200 g/d can generally be considered diarrheal. Diarrhea may be further defined as *acute* if <2 weeks, *persistent* if 2–4 weeks, and *chronic* if >4

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

DIARRHEA

DEFINITION

Diarrhea is loosely defined as passage of abnormally liquid or unformed stools at an increased frequency. For adults on a typical Western diet, stool weight >200 g/d can generally be considered diarrheal. Diarrhea may be further defined as *acute* if <2 weeks, *persistent* if 2–4 weeks, and *chronic* if >4 weeks in duration.

Two common conditions, usually associated with the passage of stool totaling <200 g/d, must be distinguished from diarrhea, because diagnostic and therapeutic algorithms differ. *Pseudodiarrhea*, or the frequent passage of small volumes of stool, is often associated with rectal urgency, tenesmus, or a feeling of incomplete evacuation and accompanies IBS or proctitis. *Fecal incontinence* is the involuntary discharge of rectal contents and is most often caused by neuromuscular disorders or structural anorectal problems. Diarrhea and urgency, especially if severe, may aggravate or cause incontinence. Pseudodiarrhea and fecal incontinence occur at prevalence rates comparable to or higher than that of chronic diarrhea and should always be considered in patients complaining of "diarrhea." Overflow diarrhea may occur in nursing home patients due to fecal impaction that is readily detectable by rectal examination. A careful history and physical examination generally allow these conditions to be discriminated from true diarrhea.

ACUTE DIARRHEA

More than 90% of cases of acute diarrhea are caused by infectious agents; these cases are often accompanied by vomiting, fever, and abdominal pain. The remaining 10% or so are caused by medications, toxic ingestions, ischemia, food indiscretions, and other conditions.

Infectious Agents

Most infectious diarrheas are acquired by fecal-oral transmission or, more commonly, via ingestion of food or water contaminated with pathogens from human or animal feces. In the immunocompetent person, the resident fecal microflora, containing >500 taxonomically distinct species, are rarely the source of diarrhea and may actually play a role in suppressing the growth of ingested pathogens. Disturbances of flora by antibiotics can lead to diarrhea by reducing the digestive function or by allowing the overgrowth of pathogens, such as *Clostridium difficile* **(Chap. 134)**. Acute infection or injury occurs when the ingested agent overwhelms or bypasses the host's mucosal immune and nonimmune (gastric acid, digestive enzymes, mucus secretion, peristalsis, and suppressive resident flora) defenses. Established clinical associations with specific enteropathogens may offer diagnostic clues. Diarrhea occasionally is an early symptom of infection such as SARS-CoV-2 and *Legionella*.

In the United States, five high-risk groups are recognized:

1. *Travelers.* Nearly 40% of tourists to endemic regions of Latin America, Africa, and Asia develop so-called traveler's diarrhea, most commonly due to enterotoxigenic or enteroaggregative *Escherichia coli* as well as to *Campylobacter*, *Shigella*, *Aeromonas*, norovirus, *Coronavirus*, and *Salmonella*. Visitors to Russia (especially St. Petersburg) may have increased risk of *Giardia*-associated diarrhea; visitors to Nepal may acquire *Cyclospora*. Campers, backpackers, and swimmers in wilderness areas may become infected with *Giardia*. Cruise ships may be affected by outbreaks of gastroenteritis caused by agents such as norovirus.

2. *Consumers of certain foods.* Diarrhea closely following food consumption at a picnic, banquet, or restaurant may suggest infection with *Salmonella*, *Campylobacter*, or *Shigella* from chicken; enterohemorrhagic *E. coli* (O157:H7) from undercooked hamburger; *Bacillus cereus* from fried rice or other reheated food; *Staphylococcus aureus* or *Salmonella* from mayonnaise or creams; *Salmonella* from eggs; *Listeria* from fresh or frozen uncooked foods, mushrooms, or dairy products; and *Vibrio* species, *Salmonella*, or acute hepatitis A from seafood, especially if raw. State departments of public health issue communications regarding domestic and foreign food-related illnesses, often identified by rapid DNA typing (PulseNet), that cause epidemics in the United States (e.g., the *Listeria* epidemic of 2020 from imported enoki mushrooms).

3. *Immunodeficient persons.* Individuals at risk for diarrhea include those with either primary immunodeficiency (e.g., IgA deficiency, common variable hypogammaglobulinemia, chronic granulomatous disease) or the much more common secondary immunodeficiency states (e.g., AIDS, senescence, pharmacologic suppression). Common enteric pathogens often cause a more severe and protracted diarrheal illness, and, particularly in persons with AIDS, opportunistic infections, such as by *Mycobacterium* species, certain viruses (cytomegalovirus, adenovirus, and herpes simplex), and protozoa (*Cryptosporidium*, *Isospora belli*, Microsporidia, and *Blastocystis hominis*) may also play a role **(Chap. 202)**. In patients with AIDS, agents transmitted venereally per rectum or by extension from vaginal infection (e.g., *Neisseria gonorrhoeae*, *Treponema pallidum*, with *Vibrio* species and *Yersinia* infections and should avoid raw fish and exposing open wounds to seawater.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR

simplex), and protozoa (*Cryptosporidium*, *Isospora belli*, microsporidia, and *Blastocystis hominis*) may also play a role (Chap. 202). In patients with AIDS, agents transmitted venereally per rectum or by extension from vaginal infection (e.g., *Neisseria gonorrhoeae*, *Treponema pallidum*, *Chlamydia*) may contribute to proctocolitis. Symptoms suggesting anorectal disease, particularly pain, may result from constipation occurring coincidentally in a person with immunodeficiency. Persons with hemochromatosis are especially prone to invasive, even fatal, enteric infections with *Vibrio* species and *Yersinia* infections and should avoid raw fish and exposing open wounds to seawater.

4. *Daycare attendees and their family members.* Infections with *Shigella*, *Giardia*, *Cryptosporidium*, rotavirus, and other agents are very common and should be considered.

5. *Institutionalized persons.* Infectious diarrhea is one of the most frequent categories of nosocomial infections in many hospitals and long-term care facilities; the causes are a variety of microorganisms but most commonly *C. difficile*. *C. difficile* can affect those with no history of antibiotic use and is often community acquired.

The pathophysiology underlying acute diarrhea by infectious agents produces specific clinical features that may also be helpful in diagnosis (Table 46-2). Profuse, watery diarrhea secondary to small-bowel hypersecretion occurs with ingestion of preformed bacterial toxins, enterotoxin-producing bacteria, and enteroadherent pathogens. Diarrhea associated with marked vomiting and minimal or no fever may occur abruptly within a few hours after ingestion of the former two types; vomiting is usually less, abdominal cramping or bloating is greater, and fever is higher with the latter. Cytotoxin-producing and invasive microorganisms all cause high fever and abdominal pain. Invasive bacteria and *Entamoeba histolytica* often cause bloody diarrhea (referred to as *dysentery*). *Yersinia* invades the terminal ileal and proximal colon mucosa and may cause especially severe abdominal pain with tenderness mimicking acute appendicitis.

TABLE 46-2

Association Between Pathobiology of Causative Agents and Clinical Features in Acute Infectious Diarrhea

PATHOBIOLOGY/AGENTS	INCUBATION PERIOD	VOMITING	ABDOMINAL PAIN	FEVER	DIARRHEA
Toxin producers					
Preformed toxin					
Bacillus cereus, *Staphylococcus aureus*, *Clostridium perfringens*	1–8 h 8–24 h	3–4+	1–2+	0–1+	3–4+, watery
Enterotoxin					
Vibrio cholerae, enterotoxigenic *Escherichia coli*, *Klebsiella pneumoniae*, *Aeromonas* species	8–72 h	2–4+	1–2+	0–1+	3–4+, watery
Enteroadherent					
Enteropathogenic and enteroadherent *E. coli*, *Giardia* organisms, cryptosporidiosis, helminths	1–8 d	0–1+	1–3+	0–2+	1–2+, watery, mushy
Cytotoxin producers					
Clostridium difficile	1–3 d	0–1+	3–4+	1–2+	1–3+, usually watery, occasionally bloody
Hemorrhagic *E. coli*	12–72 h	0–1+	3–4+	1–2+	1–3+, initially watery, quickly bloody
Invasive organisms					
Rotavirus and norovirus	1–3 d	1–3+	2–3+	3–4+	1–3+, watery

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

after ingestion of the former two types: vomiting is usually less, abdominal cramping or bloating is greater, and fever is higher with the latter.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 46-2

Association Between Pathobiology of Causative Agents and Clinical Features in Acute Infectious Diarrhea

PATHOBIOLOGY/AGENTS	INCUBATION PERIOD	VOMITING	ABDOMINAL PAIN	FEVER	DIARRHEA
Toxin producers					
Preformed toxin					
Bacillus cereus, *Staphylococcus aureus*, *Clostridium perfringens*	1–8 h 8–24 h	3–4+	1–2+	0–1+	3–4+, watery
Enterotoxin					
Vibrio cholerae, enterotoxigenic *Escherichia coli*, *Klebsiella pneumoniae*, *Aeromonas* species	8–72 h	2–4+	1–2+	0–1+	3–4+, watery
Enteroadherent					
Enteropathogenic and enteroadherent *E. coli*, *Giardia* organisms, cryptosporidiosis, helminths	1–8 d	0–1+	1–3+	0–2+	1–2+, watery, mushy
Cytotoxin producers					
Clostridium difficile	1–3 d	0–1+	3–4+	1–2+	1–3+, usually watery, occasionally bloody
Hemorrhagic *E. coli*	12–72 h	0–1+	3–4+	1–2+	1–3+, initially watery, quickly bloody
Invasive organisms					
Minimal inflammation					
Rotavirus and norovirus	1–3 d	1–3+	2–3+	3–4+	1–3+, watery
Variable inflammation					
Salmonella, *Campylobacter*, and *Aeromonas* species, *Vibrio parahaemolyticus*, *Yersinia*	12 h–11 d	0–3+	2–4+	3–4+	1–4+, watery or bloody
Severe inflammation					
Shigella species, enteroinvasive *E. coli*, *Entamoeba histolytica*	12 h–8 d	0–1+	3–4+	3–4+	1–2+, bloody

Source: Adapted from DW Powell, in T Yamada (ed): *Textbook of Gastroenterology and Hepatology*, 4th ed. Philadelphia, Lippincott Williams & Wilkins, 2003.

Finally, infectious diarrhea may be associated with systemic manifestations. Reactive arthritis (formerly known as Reiter's syndrome), arthritis, urethritis, and conjunctivitis may accompany or follow infections by *Salmonella*, *Campylobacter*, *Shigella*, and *Yersinia*. Yersiniosis may also lead to an autoimmune-type thyroiditis, pericarditis, and glomerulonephritis. Both enterohemorrhagic *E. coli* (O157:H7) and *Shigella* can lead to the *hemolytic-*

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Other Causes

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

urethritis, and conjunctivitis may accompany or follow infections by *Salmonella*, *Campylobacter*, *Shigella*, and *Yersinia*. Yersiniosis may also lead to an autoimmune-type thyroiditis, pericarditis, and glomerulonephritis. Both enterohemorrhagic *E. coli* (O157:H7) and *Shigella* can lead to the *hemolytic-uremic syndrome* with an attendant high mortality rate. The syndrome of postinfectious IBS has now been recognized as a complication of infectious diarrhea. Similarly, acute gastroenteritis may precede the diagnosis of celiac disease or Crohn's disease. Acute diarrhea can also be a major symptom of several systemic infections including *viral hepatitis*, *listeriosis*, *legionellosis*, and *toxic shock syndrome*.

Other Causes

Side effects from medications are probably the most common noninfectious causes of acute diarrhea, and etiology may be suggested by a temporal association between use and symptom onset. Although innumerable medications may produce diarrhea, some of the more frequently incriminated include antibiotics, cardiac antidysrhythmics, antihypertensives, nonsteroidal anti-inflammatory drugs (NSAIDs), certain antidepressants, chemotherapeutic agents, bronchodilators, antacids, and laxatives. Occlusive or nonocclusive ischemic colitis typically occurs in persons aged >50 years; often presents as acute lower abdominal pain preceding watery, then bloody diarrhea; and generally results in acute inflammatory changes in the sigmoid or left colon while sparing the rectum. Acute diarrhea may accompany colonic diverticulitis and graft-versus-host disease. Acute diarrhea, often associated with systemic compromise, can follow ingestion of toxins including organophosphate insecticides, amanita and other mushrooms, arsenic, and preformed toxins in seafood such as ciguatera (from algae that the fish eat) and scombroid (an excess of histamine due to inadequate refrigeration). Acute anaphylaxis to food ingestion can have a similar presentation. Conditions causing chronic diarrhea can also be confused with acute diarrhea early in their course. This confusion may occur with inflammatory bowel disease (IBD) and some of the other inflammatory chronic diarrheas that may have an abrupt rather than insidious onset and exhibit features that mimic infection.

APPROACH TO THE PATIENT

Acute Diarrhea

The decision to evaluate acute diarrhea depends on its severity and duration and on various host factors **(Fig. 46-3)**. Most episodes of acute diarrhea are mild and self-limited and do not justify the cost and potential morbidity rate of diagnostic or pharmacologic interventions. Indications for evaluation include profuse diarrhea with dehydration, grossly bloody stools, fever ≥38.5°C (≥101°F), duration >48 h without improvement, recent antibiotic use, new community outbreaks, associated severe abdominal pain in patients aged >50 years, and elderly (≥70 years) or immunocompromised patients. In some cases of moderately severe febrile diarrhea associated with fecal leukocytes (or increased fecal levels of the leukocyte proteins, such as calprotectin) or with gross blood, a diagnostic evaluation might be avoided in favor of an empirical antibiotic trial (see below).

The cornerstone of diagnosis in those suspected of severe acute infectious diarrhea is microbiologic analysis of the stool. Workup includes cultures for bacterial and viral pathogens; direct inspection for ova and parasites; and immunoassays for certain bacterial toxins (*C. difficile*), viral antigens (rotavirus), and protozoal antigens (*Giardia*, *E. histolytica*). The aforementioned clinical and epidemiologic associations may assist in focusing the evaluation. If a particular pathogen or set of possible pathogens is so implicated, either the whole panel of routine studies may not be necessary or, in some instances, special cultures may be appropriate, as for enterohemorrhagic and other types of *E. coli*, *Vibrio* species, and *Yersinia*. Molecular diagnosis of pathogens in stool can be made by identification of unique DNA sequences, and evolving microarray technologies have led to more rapid, sensitive, specific, and cost-effective diagnosis.

Persistent diarrhea is commonly due to *Giardia* **(Chap. 223)**, but additional causative organisms that should be considered include *C. difficile* (especially if antibiotics had been administered), *E. histolytica*, *Cryptosporidium*, *Campylobacter*, and others. If stool studies are unrevealing, flexible sigmoidoscopy with biopsies and upper endoscopy with duodenal aspirates and biopsies may be indicated. Brainerd diarrhea is an increasingly recognized entity characterized by an abrupt-onset diarrhea that persists for at least 4 weeks, but may last 1–3 years, and is thought to be of infectious origin. It may be associated with subtle inflammation of the distal small intestine or proximal colon.

Structural examination by sigmoidoscopy, colonoscopy, or abdominal computed tomography (CT) scanning (or other imaging approaches) may be appropriate in patients with uncharacterized persistent diarrhea to exclude IBD or as an initial approach in patients with suspected noninfectious acute diarrhea such as might be caused by ischemic colitis, diverticulitis, or partial bowel obstruction.

FIGURE 46-3

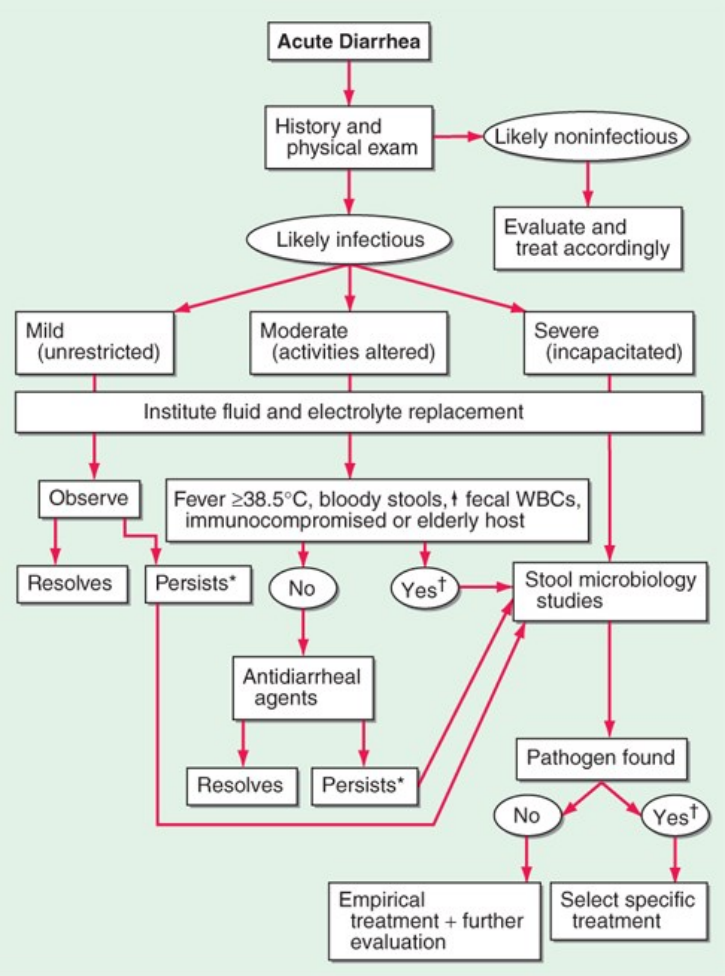
Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Algorithm for the management of acute diarrhea. Consider empirical treatment before evaluation with (*) metronidazole and with (†) quinolone. WBCs, white blood cells.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TREATMENT

Acute Diarrhea

Fluid and electrolyte replacement are of central importance to all forms of acute diarrhea. Fluid replacement alone may suffice for mild cases. Oral sugar-electrolyte solutions (iso-osmolar sport drinks or designed formulations) should be instituted promptly with severe diarrhea to limit dehydration, which is the major cause of death. Profoundly dehydrated patients, especially infants and the elderly, require IV rehydration.

In moderately severe nonfebrile and nonbloody diarrhea, antimotility and antisecretory agents such as loperamide can be useful adjuncts to control symptoms. Such agents should be avoided with febrile dysentery, which may be prolonged by them, and should be used with caution with drugs that increase levels due to cardiotoxicity. Bismuth subsalicylate may reduce symptoms of vomiting and diarrhea but should not be used to treat immunocompromised patients or those with renal impairment because of the risk of bismuth encephalopathy.

Judicious use of antibiotics is appropriate in selected instances of acute diarrhea and may reduce its severity and duration (Fig. 46-3). Many physicians treat moderately to severely ill patients with febrile dysentery empirically without diagnostic evaluation using a quinolone, such as ciprofloxacin (500 mg bid for 3–5 d). Empirical treatment can also be considered for suspected giardiasis with metronidazole (250 mg qid for 7 d). Selection of antibiotics and dosage regimens are otherwise dictated by specific pathogens, geographic patterns of resistance, and conditions found **(Chaps. 133, 161, and 165–171)**. Because of resistance to first-line treatments, newer agents such as nitazoxanide may be required for *Giardia*

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

of acquired diarrhea would be especially high, including those with immunocompromise, IBD, hemochromatosis, or gastric achlorhydria. Use of ciprofloxacin, azithromycin, or rifaximin may reduce bacterial diarrhea in such travelers by 90%, though rifaximin is not suitable for invasive disease

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Acute Diarrhea

Fluid and electrolyte replacement are of central importance to all forms of acute diarrhea. Fluid replacement alone may suffice for mild cases. Oral sugar-electrolyte solutions (iso-osmolar sport drinks or designed formulations) should be instituted promptly with severe diarrhea to limit dehydration, which is the major cause of death. Profoundly dehydrated patients, especially infants and the elderly, require IV rehydration.

In moderately severe nonfebrile and nonbloody diarrhea, antimotility and antisecretory agents such as loperamide can be useful adjuncts to control symptoms. Such agents should be avoided with febrile dysentery, which may be prolonged by them, and should be used with caution with drugs that increase levels due to cardiotoxicity. Bismuth subsalicylate may reduce symptoms of vomiting and diarrhea but should not be used to treat immunocompromised patients or those with renal impairment because of the risk of bismuth encephalopathy.

Judicious use of antibiotics is appropriate in selected instances of acute diarrhea and may reduce its severity and duration (Fig. 46-3). Many physicians treat moderately to severely ill patients with febrile dysentery empirically without diagnostic evaluation using a quinolone, such as ciprofloxacin (500 mg bid for 3–5 d). Empirical treatment can also be considered for suspected giardiasis with metronidazole (250 mg qid for 7 d). Selection of antibiotics and dosage regimens are otherwise dictated by specific pathogens, geographic patterns of resistance, and conditions found **(Chaps. 133, 161, and 165–171)**. Because of resistance to first-line treatments, newer agents such as nitazoxanide may be required for *Giardia* and *Cryptosporidium* infections. Antibiotic coverage is indicated, whether or not a causative organism is discovered, in patients who are immunocompromised, have mechanical heart valves or recent vascular grafts, or are elderly. Bismuth subsalicylate may reduce the frequency of traveler's diarrhea. Antibiotic prophylaxis is only indicated for certain patients traveling to high-risk countries in whom the likelihood or seriousness of acquired diarrhea would be especially high, including those with immunocompromise, IBD, hemochromatosis, or gastric achlorhydria. Use of ciprofloxacin, azithromycin, or rifaximin may reduce bacterial diarrhea in such travelers by 90%, though rifaximin is not suitable for invasive disease but rather as treatment for uncomplicated traveler's diarrhea. There is little role for endoscopic evaluation in most circumstances except in immunocompromised patients. Finally, physicians should be vigilant to identify if an outbreak of diarrheal illness is occurring and to alert the public health authorities promptly. This may reduce the ultimate size of the affected population.

CHRONIC DIARRHEA

Diarrhea lasting >4 weeks warrants evaluation to exclude serious underlying pathology. In contrast to acute diarrhea, most of the causes of chronic diarrhea are noninfectious. The classification of chronic diarrhea by pathophysiologic mechanism facilitates a rational approach to management, although many diseases cause diarrhea by more than one mechanism **(Table 46-3)**.

TABLE 46-3

Major Causes of Chronic Diarrhea According to Predominant Pathophysiologic Mechanism

Secretory Causes
Exogenous stimulant laxatives
Chronic ethanol ingestion
Other drugs and toxins
Endogenous laxatives (dihydroxy bile acids)
Idiopathic secretory diarrhea or bile acid diarrhea
Certain bacterial infections
Bowel resection, disease, or fistula (↓ absorption)
Hormone-producing tumors (carcinoid, VIPoma, medullary cancer of thyroid, mastocytosis, gastrinoma, colorectal villous adenoma)

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

~~Bowel resection, disease, or fistula (↓ absorption)~~
Partial bowel obstruction or fecal impaction
Hormone-producing tumors (carcinoid, VIPoma, medullary cancer of thyroid, mastocytosis, gastrinoma, colorectal villous adenoma)
Addison's disease
Congenital electrolyte absorption defects
Osmotic Causes
Osmotic laxatives (Mg^{2+}, PO_4^{-3}, SO_4^{-2})
Lactase and other disaccharide deficiencies
Nonabsorbable carbohydrates (sorbitol, lactulose, polyethylene glycol)
Gluten and FODMAP intolerance
Steatorrheal Causes
Intraluminal maldigestion (pancreatic exocrine insufficiency, bacterial overgrowth, bariatric surgery, liver disease)
Mucosal malabsorption (celiac sprue, Whipple's disease, infections, abetalipoproteinemia, ischemia, drug-induced enteropathy)
Postmucosal obstruction (1° or 2° lymphatic obstruction)
Inflammatory Causes
Idiopathic inflammatory bowel disease (Crohn's, chronic ulcerative colitis)
Lymphocytic and collagenous colitis
Immune-related mucosal disease (1° or 2° immunodeficiencies, food allergy, eosinophilic gastroenteritis, graft-versus-host disease)
Infections (invasive bacteria, viruses, and parasites, Brainerd diarrhea)
Radiation injury
Gastrointestinal malignancies
Dysmotile Causes
Irritable bowel syndrome (including postinfectious IBS)
Visceral neuromyopathies
Hyperthyroidism
Drugs (prokinetic agents)
Postvagotomy
Munchausen

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Postvagotomy
Factitial Causes
Munchausen
Eating disorders
Iatrogenic Causes
Cholecystectomy
Ileal resection
Bariatric surgery
Vagotomy, fundoplication

Abbreviations: FODMAP, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; IBS, irritable bowel syndrome.

Secretory Causes

Secretory diarrheas are due to derangements in fluid and electrolyte transport across the enterocolonic mucosa. They are characterized clinically by watery, large-volume fecal outputs that are typically painless and persist with fasting. Because there is no malabsorbed solute, stool osmolality is accounted for by normal endogenous electrolytes with no fecal osmotic gap.

MEDICATIONS

Side effects from regular ingestion of drugs and toxins are the most common secretory causes of chronic diarrhea. Hundreds of prescription and over-the-counter medications (see earlier section, “Acute Diarrhea, Other Causes”) may produce diarrhea. Surreptitious or habitual use of stimulant laxatives (e.g., senna, cascara, bisacodyl, ricinoleic acid [castor oil]) must also be considered. Chronic ethanol consumption may cause a secretory-type diarrhea due to enterocyte injury with impaired sodium and water absorption as well as rapid transit and other alterations. Inadvertent ingestion of certain environmental toxins (e.g., arsenic) may lead to chronic rather than acute forms of diarrhea. Certain bacterial infections may occasionally persist and be associated with a secretory-type diarrhea. The oral angiotensin receptor blocker olmesartan is associated with diarrhea due to sprue-like enteropathy.

BOWEL RESECTION, MUCOSAL DISEASE, OR ENTEROCOLIC FISTULA

These conditions may result in a secretory-type diarrhea because of inadequate surface for reabsorption of secreted fluids and electrolytes. Unlike other secretory diarrheas, this subset of conditions tends to worsen with eating. With disease (e.g., Crohn’s ileitis) or resection of <100 cm of terminal ileum, dihydroxy bile acids may escape absorption and stimulate colonic secretion (cholerheic diarrhea). This mechanism may contribute to so-called *idiopathic secretory diarrhea or bile acid diarrhea (BAD)*, in which bile acids are functionally malabsorbed from a normal-appearing terminal ileum. This *idiopathic bile acid malabsorption (BAM)* may account for an average of 40% of unexplained chronic diarrhea. Reduced negative feedback regulation of bile acid synthesis in hepatocytes by fibroblast growth factor 19 (FGF-19) produced by ileal enterocytes results in a degree of bile-acid synthesis that exceeds the normal capacity for ileal reabsorption, producing BAD. An alternative cause of BAD is a genetic variation in the receptor proteins (β-klotho and fibroblast growth factor 4) on the hepatocyte that normally mediate the effect of FGF-19. Dysfunction of these proteins prevents FGF-19 inhibition of hepatocyte bile acid synthesis. Another mechanism is based on genetic variation in the bile acid receptor (TGR5) in the colon, resulting in accelerated colonic transit.

Partial bowel obstruction, ostomy stricture, or fecal impaction may paradoxically lead to increased fecal output due to fluid hypersecretion.

HORMONES

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

wheezing, dyspnea, and right-sided valvular heart disease. Diarrhea is due to the release into the circulation of potent intestinal secretagogues including serotonin, histamine, prostaglandins, and various kinins. Pellagra-like skin lesions may rarely occur as the result of serotonin

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

HORMONES

Although uncommon, the classic examples of secretory diarrhea are those mediated by hormones. *Metastatic gastrointestinal carcinoid tumors* or, rarely, *primary bronchial carcinoids* may produce watery diarrhea alone or as part of the carcinoid syndrome that comprises episodic flushing, wheezing, dyspnea, and right-sided valvular heart disease. Diarrhea is due to the release into the circulation of potent intestinal secretagogues including serotonin, histamine, prostaglandins, and various kinins. Pellagra-like skin lesions may rarely occur as the result of serotonin overproduction with niacin depletion. *Gastrinoma*, one of the most common neuroendocrine tumors, most typically presents with refractory peptic ulcers, but diarrhea occurs in up to one-third of cases and may be the only clinical manifestation in 10%. While other secretagogues released with gastrin may play a role, the diarrhea most often results from fat maldigestion owing to pancreatic enzyme inactivation by low intraduodenal pH. The watery diarrhea hypokalemia achlorhydria syndrome, also called *pancreatic cholera*, is due to a non-β cell pancreatic adenoma, referred to as a *VIPoma*, that secretes VIP and a host of other peptide hormones including pancreatic polypeptide, secretin, gastrin, gastrin-inhibitory polypeptide (also called glucose-dependent insulinotropic peptide), neurotensin, calcitonin, and prostaglandins. The secretory diarrhea is often massive with stool volumes >3 L/d; daily volumes as high as 20 L have been reported. Life-threatening dehydration; neuromuscular dysfunction from associated hypokalemia, hypomagnesemia, or hypercalcemia; flushing; and hyperglycemia may accompany a VIPoma. *Medullary carcinoma of the thyroid* may present with watery diarrhea caused by calcitonin, other secretory peptides, or prostaglandins. Prominent diarrhea is often associated with metastatic disease and poor prognosis. *Systemic mastocytosis*, which may be associated with the skin lesion urticaria pigmentosa, may cause diarrhea that is either secretory and mediated by histamine or inflammatory due to intestinal infiltration by mast cells. Large *colorectal villous adenomas* may rarely be associated with a secretory diarrhea that may cause hypokalemia, can be inhibited by NSAIDs, and are apparently mediated by prostaglandins.

CONGENITAL DEFECTS IN ION ABSORPTION

Rarely, defects in specific carriers associated with ion absorption cause watery diarrhea from birth. These disorders include defective Cl^-/HCO_3^- exchange (*congenital chloridorrhea*) with alkalosis (which results from a mutated *DRA* [down-regulated in adenoma] gene) and defective Na^+/H^+ exchange (*congenital sodium diarrhea*), which results from a mutation in the *NHE3* (sodium-hydrogen exchanger) gene and results in acidosis.

Some hormone deficiencies may be associated with watery diarrhea, such as occurs with adrenocortical insufficiency (Addison's disease) that may be accompanied by skin hyperpigmentation.

Osmotic Causes

Osmotic diarrhea occurs when ingested, poorly absorbable, osmotically active solutes draw enough fluid into the lumen to exceed the reabsorptive capacity of the colon. Fecal water output increases in proportion to such a solute load. Osmotic diarrhea characteristically ceases with fasting or with discontinuation of the causative agent.

OSMOTIC LAXATIVES

Ingestion of magnesium-containing antacids, health supplements, or laxatives may induce osmotic diarrhea typified by a stool osmotic gap (>50 mosmol/L): serum osmolarity (typically 290 mosmol/kg) − (2 × [fecal sodium + potassium concentration]). Measurement of fecal osmolarity is no longer recommended because, even when measured immediately after evacuation, it may be erroneous because carbohydrates are metabolized by colonic bacteria, causing an increase in osmolarity.

CARBOHYDRATE MALABSORPTION

Carbohydrate malabsorption due to acquired or congenital defects in brush-border disaccharidases and other enzymes leads to osmotic diarrhea with a low pH. One of the most common causes of chronic diarrhea in adults is *lactase deficiency*, which affects three-fourths of nonwhites worldwide and 5–30% of persons in the United States; the total lactose load at any one time influences the symptoms experienced. Most patients learn to avoid milk products without requiring treatment with enzyme supplements. Some sugars, such as sorbitol, lactulose, or fructose, are frequently malabsorbed, and diarrhea ensues with ingestion of medications, gum, or candies sweetened with these poorly or incompletely absorbed sugars.

WHEAT AND FODMAP INTOLERANCE

Chronic diarrhea, bloating, and abdominal pain are recognized as symptoms of nonceliac gluten intolerance (which is associated with impaired intestinal or colonic barrier function) and intolerance of fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs). The

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Fat malabsorption may lead to greasy, foul-smelling, difficult-to-flush diarrhea often associated with weight loss and nutritional deficiencies due to

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Chronic diarrhea, bloating, and abdominal pain are recognized as symptoms of nonceliac gluten intolerance (which is associated with impaired intestinal or colonic barrier function) and intolerance of fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs). The latter's effects represent the interaction between the GI microbiome and the nutrients.

Steatorrheal Causes

Fat malabsorption may lead to greasy, foul-smelling, difficult-to-flush diarrhea often associated with weight loss and nutritional deficiencies due to concomitant malabsorption of amino acids and vitamins. Increased fecal output is caused by the osmotic effects of fatty acids, especially after bacterial hydroxylation, and, to a lesser extent, by the neutral fat. Quantitatively, steatorrhea is defined as stool fat exceeding the normal 7 g/d; rapid-transit diarrhea may result in fecal fat up to 14 g/d; daily fecal fat averages 15–25 g with small-intestinal diseases and is often >32 g with pancreatic exocrine insufficiency. Intraluminal maldigestion, mucosal malabsorption, or lymphatic obstruction may produce steatorrhea.

INTRALUMINAL MALDIGESTION

This condition most commonly results from pancreatic exocrine insufficiency, which occurs when >90% of pancreatic secretory function is lost. *Chronic pancreatitis*, usually a sequel of ethanol abuse, most frequently causes pancreatic insufficiency. Other causes include *cystic fibrosis*, *pancreatic duct obstruction*, and, rarely, *somatostatinoma*. Bacterial overgrowth in the small intestine may deconjugate bile acids and alter micelle formation, impairing fat digestion; it occurs with stasis from a blind-loop, small-bowel diverticulum or dysmotility and is especially likely in the elderly. Finally, cirrhosis or biliary obstruction may lead to mild steatorrhea due to deficient intraluminal bile acid concentration.

MUCOSAL MALABSORPTION

Mucosal malabsorption occurs from a variety of enteropathies, but it most commonly occurs from *celiac disease*. This gluten-sensitive enteropathy affects all ages and is characterized by villous atrophy and crypt hyperplasia in the proximal small bowel and can present with fatty diarrhea associated with multiple nutritional deficiencies of varying severity. Celiac disease is much more frequent than previously thought; it affects ~1% of the population, frequently presents without steatorrhea, can mimic IBS, and has many other GI and extraintestinal manifestations. *Tropical sprue* may produce a similar histologic and clinical syndrome but occurs in residents of or travelers to tropical climates; abrupt onset and response to antibiotics suggest an infectious etiology. *Whipple's disease*, due to the bacillus *Tropheryma whipplei* and histiocytic infiltration of the small-bowel mucosa, is a less common cause of steatorrhea that most typically occurs in young or middle-aged men; it is frequently associated with arthralgias, fever, lymphadenopathy, and extreme fatigue, and it may affect the CNS and endocardium. A similar clinical and histologic picture results from *Mycobacterium avium-intracellulare* infection in patients with AIDS. *Abetalipoproteinemia* is a rare defect of chylomicron formation and fat malabsorption in children, associated with acanthocytic erythrocytes, ataxia, and retinitis pigmentosa. Several other conditions may cause mucosal malabsorption including infections, especially with protozoa such as *Giardia*, numerous medications (e.g., olmesartan, mycophenolate mofetil, colchicine, cholestyramine, neomycin), idiopathic enteropathies, amyloidosis, and chronic ischemia.

POSTMUCOSAL LYMPHATIC OBSTRUCTION

The pathophysiology of this condition, which is due to the rare *congenital intestinal lymphangiectasia* or to *acquired lymphatic obstruction* secondary to trauma, tumor, cardiac disease, or infection, leads to the unique constellation of fat malabsorption with enteric losses of protein (often causing edema) and lymphocytopenia. Carbohydrate and amino acid absorption are preserved.

Inflammatory Causes

Inflammatory diarrheas are generally accompanied by pain, fever, bleeding, or other manifestations of inflammation. The mechanism of diarrhea may not only be exudation but, depending on lesion site, may include fat malabsorption, disrupted fluid/electrolyte absorption, and hypersecretion or hypermotility from release of cytokines and other inflammatory mediators. The unifying feature on stool analysis is the presence of leukocytes or leukocyte-derived proteins such as calprotectin. With severe inflammation, exudative protein loss can lead to anasarca (generalized edema). Any middle-aged or older person with chronic inflammatory-type diarrhea, especially with blood, should be carefully evaluated to exclude a colorectal tumor.

IDIOPATHIC INFLAMMATORY BOWEL DISEASE

The illnesses in this category, which include *Crohn's disease* and *chronic ulcerative colitis*, are among the most common organic causes of chronic diarrhea in adults and range in severity from mild to fulminant and life-threatening. They may be associated with uveitis, polyarthralgias, cholestatic

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

histologic diagnosis. It may coexist with symptoms suggesting IBS or with celiac sprue or drug-induced enteropathy. It typically responds well to anti-inflammatory drugs (e.g., bismuth), the opioid agonist loperamide, or budesonide.

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

The illnesses in this category, which include *Crohn's disease* and *chronic ulcerative colitis*, are among the most common organic causes of chronic diarrhea in adults and range in severity from mild to fulminant and life-threatening. They may be associated with uveitis, polyarthralgias, cholestatic liver disease (primary sclerosing cholangitis), and skin lesions (erythema nodosum, pyoderma gangrenosum). *Microscopic colitis*, including both lymphocytic and *collagenous colitis*, is an increasingly recognized cause of chronic watery diarrhea, especially in middle-aged women and those on NSAIDs, statins, proton pump inhibitors (PPIs), and selective serotonin reuptake inhibitors (SSRIs); biopsy of a normal-appearing colon is required for histologic diagnosis. It may coexist with symptoms suggesting IBS or with celiac sprue or drug-induced enteropathy. It typically responds well to anti-inflammatory drugs (e.g., bismuth), the opioid agonist loperamide, or budesonide.

PRIMARY OR SECONDARY FORMS OF IMMUNODEFICIENCY

Immunodeficiency may lead to prolonged infectious diarrhea. With selective IgA deficiency or common variable *hypogammaglobulinemia*, diarrhea is particularly prevalent and often the result of giardiasis, bacterial overgrowth, or sprue.

EOSINOPHILIC GASTROENTERITIS

Eosinophil infiltration of the mucosa, muscularis, or serosa at any level of the GI tract may cause diarrhea, pain, vomiting, or ascites. Affected patients often have an atopic history, Charcot-Leyden crystals due to extruded eosinophil contents may be seen on microscopic inspection of stool, and peripheral eosinophilia is present in 50–75% of patients. While hypersensitivity to certain foods occurs in adults, true food allergy causing chronic diarrhea is rare.

OTHER CAUSES

Chronic inflammatory diarrhea may be caused by *radiation enterocolitis*, *chronic graft-versus-host disease*, autoimmune or idiopathic enteropathies, *Behçet's syndrome*, and *Cronkhite-Canada syndrome*, among others.

Dysmotility Causes

Rapid transit may accompany many diarrheas as a secondary or contributing phenomenon, but primary dysmotility is an unusual etiology of true diarrhea. Stool features often suggest a secretory diarrhea, but mild steatorrhea of up to 14 g of fat per day can be produced by maldigestion from rapid transit alone. *Hyperthyroidism*, *carcinoid syndrome*, and certain drugs (e.g., prostaglandins, prokinetic agents) may produce hypermotility with resultant diarrhea. Primary visceral neuromyopathies or idiopathic acquired intestinal pseudoobstruction may lead to stasis with secondary bacterial overgrowth causing diarrhea. *Diabetic diarrhea*, often accompanied by peripheral and generalized autonomic neuropathies, may occur in part because of intestinal dysmotility.

The exceedingly common IBS (10% point prevalence, 1–2% per year incidence) is characterized by disturbed intestinal and colonic motor and sensory responses to various stimuli. Symptoms of stool frequency typically cease at night, alternate with periods of constipation, are accompanied by abdominal pain relieved with defecation, and rarely result in weight loss.

Factitial Causes

Factitial diarrhea accounts for up to 15% of unexplained diarrheas referred to tertiary care centers. Either as a form of *Munchausen syndrome* (deception or self-injury for secondary gain) or *eating disorders*, some patients covertly self-administer laxatives alone or in combination with other medications (e.g., diuretics) or surreptitiously add water or urine to stool sent for analysis. Such patients are typically women, often with histories of psychiatric illness, and disproportionately from careers in health care. Hypotension and hypokalemia are common co-presenting features. The evaluation of such patients may be difficult: contamination of the stool with water or urine is suggested by very low or high stool osmolarity, respectively. Such patients often deny this possibility when confronted, but they do benefit from psychiatric counseling when they acknowledge their behavior.

APPROACH TO THE PATIENT

CHRONIC DIARRHEA

The laboratory tools available to evaluate the very common problem of chronic diarrhea are extensive, and many are costly and invasive. As such, the diagnostic evaluation must be rationally directed by a careful history, including medications, and physical examination **(Fig. 46-4)**. When this

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

aggravating (especially diet) and relieving factors, and stool characteristics of their diarrhea. The presence or absence of fecal incontinence, fever, weight loss, pain, certain exposures (travel, medications, contacts with diarrhea), and common extraintestinal manifestations (skin changes,

the diagnostic evaluation must be rationally directed by a careful history, including medications, and physical examination **(Fig. 46-4)**. When this strategy is unrevealing, simple triage tests are often warranted to direct the choice of more complex investigations (Fig. 46-4). The history, physical examination **(Table 46-4)**, and routine blood studies should attempt to characterize the mechanism of diarrhea, identify diagnostically helpful associations, and assess the patient's fluid/electrolyte and nutritional status. Patients should be questioned about the onset, duration, pattern, aggravating (especially diet) and relieving factors, and stool characteristics of their diarrhea. The presence or absence of fecal incontinence, fever, weight loss, pain, certain exposures (travel, medications, contacts with diarrhea), and common extraintestinal manifestations (skin changes, arthralgias, oral aphthous ulcers) should be noted. A family history of IBD or celiac disease may indicate those possibilities. Physical findings may offer clues such as a thyroid mass, wheezing, heart murmurs, edema, hepatomegaly, abdominal masses, lymphadenopathy, mucocutaneous abnormalities, perianal fistulas, or anal sphincter laxity. Peripheral blood leukocytosis, elevated sedimentation rate, or C-reactive protein suggests inflammation; anemia reflects blood loss or nutritional deficiencies; or eosinophilia may occur with parasitoses, neoplasia, collagen-vascular disease, allergy, or eosinophilic gastroenteritis. Blood chemistries may demonstrate electrolyte, hepatic, or other metabolic disturbances. Measuring IgA tissue transglutaminase antibodies may help detect celiac disease. Bile acid diarrhea is confirmed by a scintigraphic radiolabeled bile acid retention test; however, this is not available in many countries. Alternative approaches are a screening blood test (serum C4 or FGF-19), measurement of fecal bile acids, or a therapeutic trial with a bile acid sequestrant (e.g., cholestyramine, colestipol or colesevelam).

A therapeutic trial is often appropriate, definitive, and highly cost-effective when a specific diagnosis is suggested on the initial physician encounter. For example, chronic watery diarrhea, which ceases with fasting in an otherwise healthy young adult, may justify a trial of a lactose-restricted diet; bloating and diarrhea persisting since a mountain backpacking trip may warrant a trial of metronidazole for likely giardiasis; and postprandial diarrhea persisting following resection of terminal ileum might be due to bile acid malabsorption and be treated with cholestyramine, colestipol, or colesevelam before further evaluation. Persistent symptoms require additional investigation.

Certain diagnoses may be suggested on the initial encounter (e.g., idiopathic IBD); however, additional focused evaluations may be necessary to confirm the diagnosis and characterize the severity or extent of disease so that treatment can be best guided. Patients suspected of having IBS should be initially evaluated with flexible sigmoidoscopy with colorectal biopsies to exclude IBD, or particularly microscopic colitis, which is clinically indistinguishable from IBS with diarrhea or functional diarrhea; those with normal findings might be reassured and, as indicated, treated empirically with antispasmodics, antidiarrheals, or antidepressants (e.g., tricyclic agents). Any patient who presents with chronic diarrhea and hematochezia should be evaluated with stool microbiologic studies and colonoscopy.

In an estimated two-thirds of cases, the cause for chronic diarrhea remains unclear after the initial encounter, and further testing is required. Quantitative stool collection and analyses can yield important objective data that may establish a diagnosis or characterize the type of diarrhea as a triage for focused additional studies (Fig. 46-4). If stool weight is >200 g/d, additional stool analyses should be performed that might include electrolyte concentration, pH, occult blood testing, leukocyte inspection (or leukocyte protein assay), fat quantitation, and laxative screens.

For secretory diarrheas (watery, normal osmotic gap), possible medication-related side effects or surreptitious laxative use should be reconsidered. Microbiologic studies should be done including fecal bacterial cultures (including media for *Aeromonas* and *Plesiomonas*), inspection for ova and parasites, and *Giardia* antigen assay (the most sensitive test for giardiasis). Small-bowel bacterial overgrowth can be excluded by intestinal aspirates with quantitative cultures or with glucose or lactulose breath tests involving measurement of breath hydrogen, methane, or other metabolite. However, interpretation of these breath tests may be confounded by disturbances of intestinal transit. Upper endoscopy and colonoscopy with biopsies and small-bowel x-rays (formerly barium, but increasingly CT with enterography or magnetic resonance with enteroclysis) are helpful to rule out structural or occult inflammatory disease. When suggested by history or other findings, screens for peptide hormones should be pursued (e.g., serum gastrin, VIP, calcitonin, thyroid hormone/thyroid-stimulating hormone, urinary 5-hydroxyindolacetic acid, histamine).

Further evaluation of osmotic diarrhea should include tests for lactose intolerance and magnesium ingestion, the two most common causes. Low fecal pH suggests carbohydrate malabsorption; lactose malabsorption can be confirmed by lactose breath testing or by a therapeutic trial with lactose exclusion and observation of the effect of lactose challenge (e.g., a liter of milk). Lactase determination on small-bowel biopsy is not generally available. If fecal magnesium or laxative levels are elevated, inadvertent or surreptitious ingestion should be considered and psychiatric help should be sought.

For those with proven fatty diarrhea, endoscopy with small-bowel biopsy (including aspiration for quantitative cultures, if available) should be performed; if this procedure is unrevealing, a small-bowel radiograph is often an appropriate next step. If small-bowel studies are negative or if pancreatic disease is suspected, pancreatic exocrine insufficiency should be excluded with direct tests, such as the secretin-cholecystokinin stimulation test or a variation that could be performed endoscopically. In general, indirect tests such as assay of fecal elastase or chymotrypsin activity or a bentiromide test have fallen out of favor because of low sensitivity and specificity.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

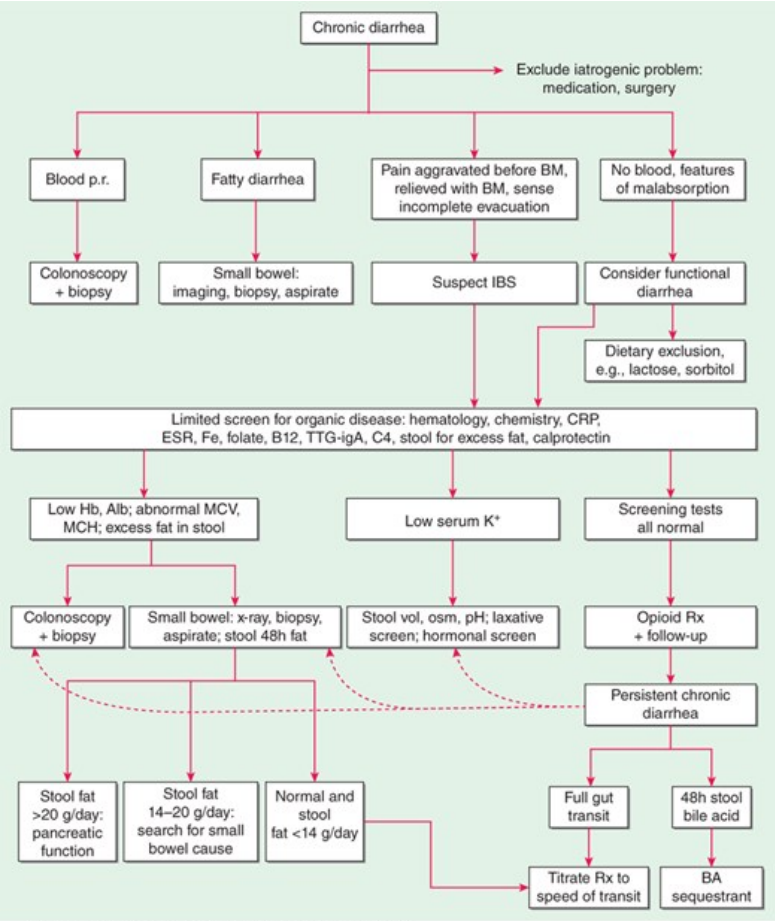
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

stimulation test or a variation that could be performed endoscopically. In general, indirect tests such as assay of fecal elastase or chymotrypsin activity or a bentiromide test have fallen out of favor because of low sensitivity and specificity.

Chronic inflammatory-type diarrheas should be suspected by the presence of blood or leukocytes in the stool. Such findings warrant stool cultures; inspection for ova and parasites; *C. difficile* toxin assay; colonoscopy with biopsies; and, if indicated, small-bowel imaging studies.

FIGURE 46-4

Algorithm for management of chronic diarrhea. Patients undergo an initial evaluation based on different symptom presentations, leading to selection of patients for imaging, biopsy analysis, and limited screens for organic diseases. Alb, albumin; BA, bile acid; BM, bowel movement; C4, 7 α-hydroxy-4-cholesten-3-one; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; Hb, hemoglobin; Hx, history; IBS, irritable bowel syndrome; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume; osm, osmolality; p.r., per rectum; Rx, treatment; TTG, tissue transglutaminase. *(Reproduced with permission from M Camilleri, JH Sellin, KE Barrett: Pathophysiology, evaluation, and management of chronic watery diarrhea, Gastroenterology 152:515, 2017.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 46-4

Physical Examination in Patients with Chronic Diarrhea

1. Are there general features to suggest malabsorption or inflammatory bowel disease (IBD) such as anemia, dermatitis herpetiformis, edema, or clubbing?
2. Are there features to suggest underlying autonomic neuropathy or collagen-vascular disease in the pupils, orthostasis, skin, hands, or joints?
3. Is there an abdominal mass or tenderness?
4. Are there any abnormalities of rectal mucosa, rectal defects, or altered anal sphincter functions?

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

TABLE 46-4

Physical Examination in Patients with Chronic Diarrhea

1. Are there general features to suggest malabsorption or inflammatory bowel disease (IBD) such as anemia, dermatitis herpetiformis, edema, or clubbing?
2. Are there features to suggest underlying autonomic neuropathy or collagen-vascular disease in the pupils, orthostasis, skin, hands, or joints?
3. Is there an abdominal mass or tenderness?
4. Are there any abnormalities of rectal mucosa, rectal defects, or altered anal sphincter functions?
5. Are there any mucocutaneous manifestations of systemic disease such as dermatitis herpetiformis (celiac disease), erythema nodosum (ulcerative colitis), flushing (carcinoid), or oral ulcers for IBD or celiac disease?

TREATMENT

Chronic Diarrhea

Treatment of chronic diarrhea depends on the specific etiology and may be curative, suppressive, or empirical. If the cause can be eradicated, treatment is curative as with resection of a colorectal cancer, antibiotic administration for Whipple's disease or tropical sprue, or discontinuation of a drug. For many chronic conditions, diarrhea can be controlled by suppression of the underlying mechanism. Examples include elimination of dietary lactose for lactase deficiency or gluten for celiac sprue, use of glucocorticoids or other anti-inflammatory agents for idiopathic IBDs, bile acid sequestrants for bile acid malabsorption, PPIs for the gastric hypersecretion of gastrinomas, somatostatin analogues such as octreotide for malignant carcinoid syndrome, prostaglandin inhibitors such as indomethacin for medullary carcinoma of the thyroid, and pancreatic enzyme replacement for pancreatic insufficiency. When the specific cause or mechanism of chronic diarrhea evades diagnosis, empirical therapy may be beneficial. Mild opiates, such as diphenoxylate or loperamide, are often helpful in mild or moderate watery diarrhea. For those with more severe diarrhea, codeine or tincture of opium may be beneficial. Such antimotility agents should be avoided with severe IBD, because toxic megacolon may be precipitated. Clonidine, an α_2-adrenergic agonist, may allow control of diabetic diarrhea, although the medication may be poorly tolerated because it causes postural hypotension. The 5-HT_3 receptor antagonists (e.g., alosetron, ondansetron) may relieve diarrhea and urgency in patients with IBS diarrhea. Other medications approved for the treatment of diarrhea associated with IBS are the nonabsorbed antibiotic, rifaximin, and the mixed μ-opioid receptor (OR) and κ-OR agonist and δ-OR antagonist, eluxadoline. The latter may induce sphincter of Oddi spasm and subsequent acute pancreatitis, usually in patients with prior cholecystectomy. For all patients with chronic diarrhea, fluid and electrolyte repletion is an important component of management (see "Acute Diarrhea," earlier). Replacement of fat-soluble vitamins may also be necessary in patients with chronic steatorrhea.

CONSTIPATION

DEFINITION

Constipation is a common complaint in clinical practice and usually refers to persistent, difficult, infrequent, or seemingly incomplete defecation. Because of the wide range of normal bowel habits, constipation is difficult to define precisely. Most persons have at least three bowel movements per week; however, low stool frequency alone is not the sole criterion for the diagnosis of constipation. Many constipated patients have a normal frequency of defecation but complain of excessive straining, hard stools, lower abdominal fullness, or a sense of incomplete evacuation. The individual patient's symptoms must be analyzed in detail to ascertain what is meant by "constipation" or "difficulty" with defecation.

Stool form and consistency are well correlated with the time elapsed from the preceding defecation. Hard, pellety stools occur with slow transit, whereas loose, watery stools are associated with rapid transit. Both small pellety or very large stools are more difficult to expel than normal stools.

The perception of hard stools or excessive straining is more difficult to assess objectively, and the need for enemas or digital disimpaction is a clinically

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

irritation; and some adults habitually ignore or delay the call to have a bowel movement.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

The perception of hard stools or excessive straining is more difficult to assess objectively, and the need for enemas or digital disimpaction is a clinically useful way to corroborate the patient's perceptions of difficult defecation.

Psychosocial or cultural factors may also be important. A person whose parents attached great importance to daily defecation will become greatly concerned when he or she misses a daily bowel movement; some children withhold stool to gain attention or because of fear of pain from anal irritation; and some adults habitually ignore or delay the call to have a bowel movement.

CAUSES

Pathophysiologically, chronic constipation generally results from inadequate fiber or fluid intake or from disordered colonic transit or anorectal function. These result from neurogastroenterologic disturbance, certain drugs, advancing age, or in association with a large number of systemic diseases that affect the GI tract **(Table 46-5)**. Constipation of recent onset may be a symptom of significant organic disease such as tumor, anorectal irritation, or stricture. In *idiopathic constipation*, a subset of patients exhibits delayed emptying of the ascending and transverse colon with prolongation of transit (often in the proximal colon) and a reduced frequency of propulsive HAPCs. *Outlet obstruction to defecation* (also called *evacuation disorders*) accounts for about a quarter of cases presenting with constipation in tertiary care and may cause delayed colonic transit, which is usually corrected by biofeedback retraining of the disordered defecation. Constipation of any cause may be exacerbated by hospitalization or chronic illnesses that lead to physical or mental impairment and result in inactivity or physical immobility.

TABLE 46-5

Causes of Constipation in Adults

TYPES OF CONSTIPATION AND CAUSES	EXAMPLES
Recent Onset	
Colonic obstruction	Neoplasm; stricture: ischemic, diverticular, inflammatory
Anal sphincter spasm	Anal fissure, painful hemorrhoids
Medications	
Chronic	
Irritable bowel syndrome	Constipation-predominant, alternating
Medications	Ca^{2+} blockers, antidepressants
Colonic pseudoobstruction	Slow-transit constipation, megacolon (rare Hirschsprung's, Chagas' diseases)
Disorders of rectal evacuation	Pelvic floor dysfunction; anismus; descending perineum syndrome; rectal mucosal prolapse; rectocele
Endocrinopathies	Hypothyroidism, hypercalcemia, pregnancy
Psychiatric disorders	Depression, eating disorders, drugs
Neurologic disease	Parkinsonism, multiple sclerosis, spinal cord injury
Generalized muscle disease	Progressive systemic sclerosis

APPROACH TO THE PATIENT

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

three bowel movements per week), consistency (lumpy/hard), excessive straining, prolonged defecation time, or need to support the perineum or digitate the anorectum to facilitate stool evacuation. These latter items identified in the history suggest the presence of a rectal evacuation disorder.

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

APPROACH TO THE PATIENT

Constipation

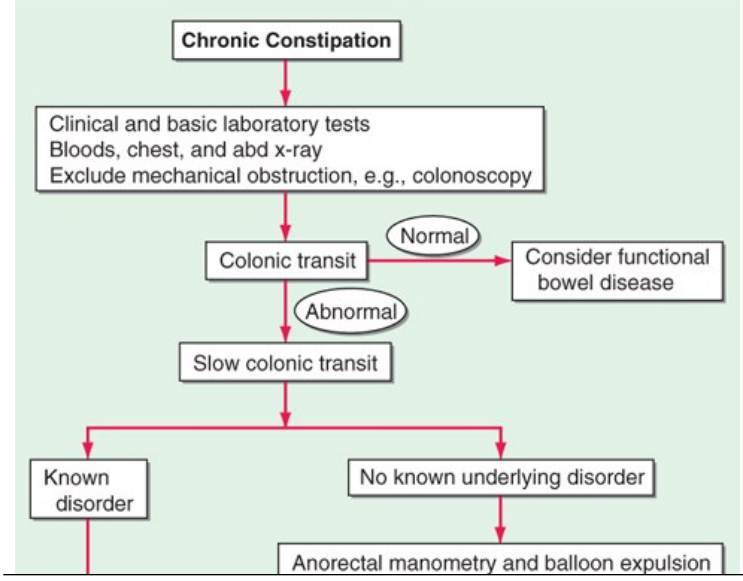
A careful history should explore the patient's symptoms and confirm whether she or he is indeed constipated based on frequency (e.g., fewer than three bowel movements per week), consistency (lumpy/hard), excessive straining, prolonged defecation time, or need to support the perineum or digitate the anorectum to facilitate stool evacuation. These latter items identified in the history suggest the presence of a rectal evacuation disorder. In the vast majority of cases (probably >90%), there is no underlying cause (e.g., cancer, depression, or hypothyroidism), and constipation responds to ample hydration, exercise, and supplementation of dietary fiber (15–25 g/d). A good diet and medication history and attention to psychosocial issues are key. Physical examination and, particularly, rectal examination are mandatory and should exclude fecal impaction and most of the important diseases that present with constipation and possibly indicate features suggesting an evacuation disorder (e.g., high anal sphincter tone, failure of perineal descent, or paradoxical puborectalis contraction or puborectalis tenderness during straining to simulate stool evacuation).

The presence of weight loss, rectal bleeding, or anemia with constipation mandates either flexible sigmoidoscopy plus barium enema or colonoscopy alone, particularly in patients aged >40 years, to exclude structural diseases such as cancer or strictures. Colonoscopy alone is most cost-effective in this setting because it provides an opportunity to biopsy mucosal lesions, perform polypectomy, or dilate strictures. Barium enema has advantages over colonoscopy in the patient with isolated constipation because it is less costly and identifies colonic dilation and all significant mucosal lesions or strictures that are likely to present with constipation. Melanosis coli, or pigmentation of the colon mucosa, indicates the use of anthraquinone laxatives such as cascara or senna; however, this is usually apparent from a careful history. An unexpected disorder such as megacolon or cathartic colon may also be detected by colonic radiographs. Measurement of serum calcium, potassium, and thyroid-stimulating hormone levels will identify rare patients with metabolic disorders.

Patients with more troublesome constipation may not respond to fiber alone and may be helped by a bowel-training regimen, which involves taking an osmotic laxative (e.g., magnesium salts, lactulose, sorbitol, polyethylene glycol) and evacuating with enema or suppository (e.g., glycerin or bisacodyl) as needed. After breakfast, a distraction-free 15–20 min on the toilet without straining is encouraged. Excessive straining may lead to development of hemorrhoids and, if there is weakness of the pelvic floor or injury to the pudendal nerve, may result in obstructed defecation from descending perineum syndrome several years later. Those few who do not benefit from the simple measures delineated above or require long-term treatment or fail to respond to potent laxatives should undergo further investigation **(Fig. 46-5)**. Novel agents that induce secretion (e.g., lubiprostone, a chloride channel activator, or linaclotide, a guanylate cyclase C agonist that activates chloride secretion) are also available.

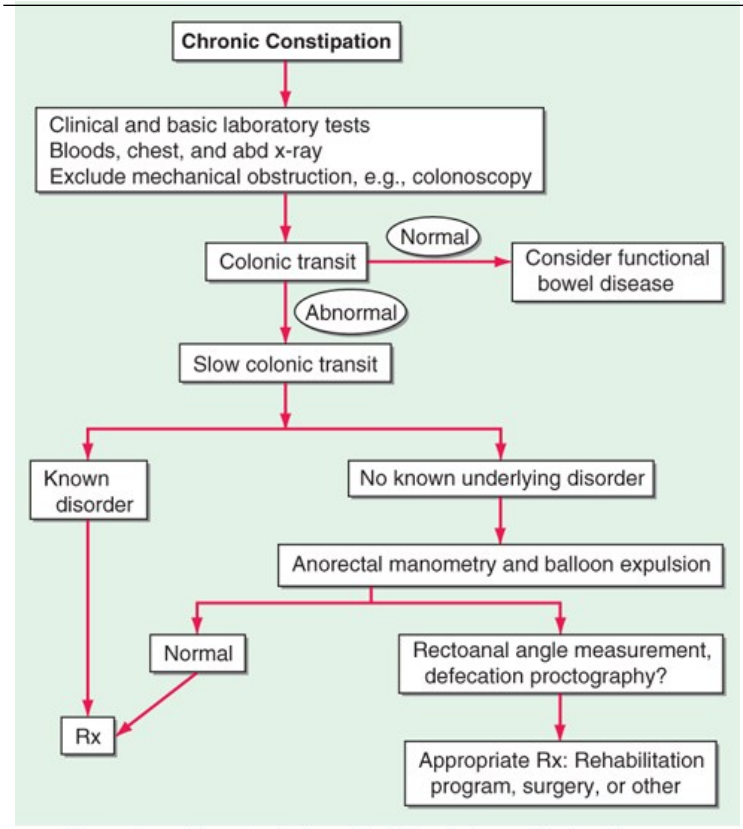
FIGURE 46-5

Algorithm for the management of constipation. abd, abdominal; Rx, treatment.



Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Algorithm for the management of constipation. abd, abdominal; Rx, treatment.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

INVESTIGATION OF SEVERE CONSTIPATION

A small minority (probably <5%) of patients have severe or "intractable" constipation; about 25% have evacuation disorders. These are the patients most likely to require evaluation by gastroenterologists or in referral centers. Further observation of the patient may occasionally reveal a previously unrecognized cause, such as an evacuation disorder, laxative abuse, malingering, or psychological disorder. In these patients, evaluations of the physiologic function of the colon and pelvic floor and of psychological status aid in the rational choice of treatment. Even among these highly selected patients with severe constipation, a cause can be identified in only about one-third of tertiary referral patients, with the others being diagnosed with normal transit constipation. Since evacuation disorders also retard colonic transit through the left colon or the entire colon, anorectal and pelvic floor testing should precede transit measurements if there is clinical suspicion of an evacuation disorder. If an evacuation disorder is identified on testing, colonic transit may be unnecessary.

Measurement of Colonic Transit

Radiopaque marker transit tests are easy, repeatable, generally safe, inexpensive, reliable, and highly applicable in evaluating constipated patients in clinical practice. Several validated methods are very simple. For example, radiopaque markers are ingested; an abdominal flat film taken 5 days later should indicate passage of 80% of the markers out of the colon without the use of laxatives or enemas. This test does not provide useful information about the transit profile of the stomach and small bowel. An alternative approach involves ingestion of 24 radiopaque markers on 3 successive days and an abdominal radiograph on the fourth day. The number of markers counted in the radiograph is an estimate of the colonic transit in hours. The collection of gas in the rectum between the level of the ischial spines and the lower border of the sacroiliac joints may suggest the presence of a rectal evacuation disorder as the cause of constipation.

Radioscintigraphy with a delayed-release capsule containing radiolabeled particles has been used to noninvasively characterize normal, accelerated, or delayed colonic function over 24–48 h with low radiation exposure. This approach simultaneously assesses gastric, small bowel (which may be important in ~20% of patients with delayed colonic transit because they reflect a more generalized GI motility disorder), and colonic transit. The disadvantages are the greater cost and the need for specific materials prepared in a nuclear medicine laboratory.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Pelvic floor dysfunction is suggested by the inability to evacuate the rectum, a feeling of persistent rectal fullness, rectal pain, the need to extract stool from the rectum digitally, application of pressure on the posterior wall of the vagina, support of the perineum during straining, and excessive straining

important in ~20% of patients with delayed colonic transit because they reflect a more generalized GI motility disorder), and colonic transit. The disadvantages are the greater cost and the need for specific materials prepared in a nuclear medicine laboratory.

Anorectal and Pelvic Floor Tests

Pelvic floor dysfunction is suggested by the inability to evacuate the rectum, a feeling of persistent rectal fullness, rectal pain, the need to extract stool from the rectum digitally, application of pressure on the posterior wall of the vagina, support of the perineum during straining, and excessive straining. These significant symptoms should be contrasted with the simple sense of incomplete rectal evacuation, which is common in IBS.

Formal psychological evaluation may identify eating disorders, “control issues,” depression, or posttraumatic stress disorders that may respond to cognitive or other intervention and may be important in restoring quality of life to patients who might present with chronic constipation.

A simple clinical test in the office to document a nonrelaxing puborectalis muscle is to have the patient strain to expel the index finger during a digital rectal examination. Motion of the puborectalis posteriorly during straining indicates proper coordination of the pelvic floor muscles. Motion anteriorly with paradoxical contraction or limited perineal descent (<1.5 cm) during simulated evacuation indicates pelvic floor dysfunction.

Measurement of perineal descent is relatively easy to gauge clinically by placing the patient in the left decubitus position and watching the perineum to detect inadequate descent (<1.5 cm, a sign of pelvic floor dysfunction) or perineal ballooning during straining relative to bony landmarks (>4 cm, suggesting excessive perineal descent).

A useful overall test of evacuation is the balloon expulsion test. A balloon-tipped urinary catheter is placed and inflated with 50 mL of water. Normally, a patient can expel it while seated on a toilet or in the left lateral decubitus position. In the lateral position, the weight needed to facilitate expulsion of the balloon is determined; normally, expulsion occurs with <200 g added or unaided within 1 minute.

Anorectal manometry, when used in the evaluation of patients with severe constipation, may find an excessively high resting (>80 mmHg) or squeeze anal sphincter tone, suggesting anismus (anal sphincter spasm). This test also identifies rare syndromes, such as adult Hirschsprung’s disease, by the absence of the rectoanal inhibitory reflex.

Defecography (a dynamic barium enema including lateral views obtained during barium expulsion or a magnetic resonance defecogram) reveals “soft abnormalities” in many patients; the most relevant findings are the measured changes in rectoanal angle, anatomic defects of the rectum such as internal mucosal prolapse, and enteroceles or rectoceles. Surgically remediable conditions are identified in only a few patients. These include severe, whole-thickness intussusception with complete outlet obstruction due to funnel-shaped plugging at the anal canal or an extremely large rectocele that fills preferentially during attempts at defecation instead of expulsion of the barium through the anus. In summary, defecography requires an interested and experienced radiologist, and abnormalities are not pathognomonic for pelvic floor dysfunction. The most common cause of outlet obstruction is failure of the puborectalis muscle to relax; this is not identified by barium defecography but can be demonstrated by magnetic resonance defecography, which provides more information about the structure and function of the pelvic floor, distal colorectum, and anal sphincters.

Neurologic testing (electromyography) is more helpful in the evaluation of patients with incontinence than of those with symptoms suggesting obstructed defecation. The absence of neurologic signs in the lower extremities suggests that any documented denervation of the puborectalis results from pelvic (e.g., obstetric) injury or from stretching of the pudendal nerve by chronic, long-standing straining. Constipation is common among patients with spinal cord injuries, neurologic diseases such as Parkinson’s disease, multiple sclerosis, and diabetic neuropathy.

Spinal-evoked responses during electrical rectal stimulation or stimulation of external anal sphincter contraction by applying magnetic stimulation over the lumbosacral cord identify patients with limited sacral neuropathies with sufficient residual nerve conduction to attempt biofeedback training.

In summary, a balloon expulsion test is an important screening test for anorectal dysfunction. Rarely, an anatomic evaluation of the rectum or anal sphincters and an assessment of pelvic floor relaxation are the tools for evaluating patients in whom obstructed defecation is suspected and is associated with symptoms of rectal mucosal prolapse, pressure of the posterior wall of the vagina to facilitate defecation (suggestive of anterior rectocele), or prior pelvic surgery that may be complicated by enterocele.

TREATMENT

Constipation

After the cause of constipation is characterized, a treatment decision can be made. Slow-transit constipation requires aggressive medical or surgical

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

require a dedicated bowel regimen that often includes rectal stimulation, enema therapy, and carefully timed laxative therapy.

Patients with constipation are treated with bulk (fiber, psyllium), osmotic (milk of magnesia, lactulose, polyethylene glycol), secretory (lubiprostone

Access Provided by:
SILVERCHAIR

TREATMENT

Constipation

After the cause of constipation is characterized, a treatment decision can be made. Slow-transit constipation requires aggressive medical or surgical treatment; anismus or pelvic floor dysfunction usually responds to biofeedback management (Fig. 46-5). The remaining ~60% of patients with constipation have normal colonic transit and can be treated symptomatically. Patients with spinal cord injuries or other neurologic disorders require a dedicated bowel regimen that often includes rectal stimulation, enema therapy, and carefully timed laxative therapy.

Patients with constipation are treated with bulk (fiber, psyllium), osmotic (milk of magnesia, lactulose, polyethylene glycol), secretory (lubiprostone, linaclotide, plecanatide, tenapanor), and prokinetic or stimulant laxatives (including diphenyl methanes such as bisacodyl and sodium picosulfate and 5-HT_4 agonists prucalopride and tegaserod). If a 3- to 6-month trial of medical therapies fails, unassociated with obstructed defecation, the patient should be considered for laparoscopic colectomy with ileorectostomy; however, this should not be undertaken for pain or if there is continued evidence of an evacuation disorder or a generalized GI dysmotility. Referral to a specialized center for further tests of colonic motor function is warranted. The decision to resort to surgery is facilitated by the presence of megacolon and megarectum. The complications after surgery include small-bowel obstruction (11%) and fecal soiling, particularly at night during the first postoperative year. Frequency of defecation is 3–8 per day during the first year, dropping to 1–3 per day from the second year after surgery.

Patients who have a combined (evacuation and transit/motility) disorder should first pursue pelvic floor retraining (biofeedback and muscle relaxation), psychological counseling, and dietetic advice. If symptoms are intractable despite biofeedback and optimized medical therapy, colectomy and ileorectostomy could be considered as long as the evacuation disorder is resolved and optimized medical therapy is unsuccessful. In patients with pelvic floor dysfunction alone, biofeedback training has a 70–80% success rate, measured by the acquisition of comfortable stool habits. Attempts to manage pelvic floor dysfunction with operations (internal anal sphincter or puborectalis muscle division) or injections with botulinum toxin have achieved only mediocre success and have been largely abandoned.

FURTHER READING

Assi R et al: Sexually transmitted infections of the anus and rectum. World J Gastroenterol 20:15262, 2014. [PubMed: 25386074]

Bharucha AE, Rao SS: An update on anorectal disorders for gastroenterologists. Gastroenterology 146:37, 2014. [PubMed: 24211860]

Bharucha AE et al: American Gastroenterological Association technical review on constipation. Gastroenterology 144:218, 2013. [PubMed: 23261065]

Boeckxstaens G et al: Fundamentals of neurogastroenterology: Physiology/motility—sensation. Gastroenterology 150:1292, 2016.

Camilleri M et al: Chronic constipation. Nat Rev Dis Primers 3:17095, 2017. [PubMed: 29239347]

Camilleri M et al: Pathophysiology, evaluation, and management of chronic watery diarrhea. Gastroenterology 152:515, 2017. [PubMed: 27773805]

Peery AF et al: Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: Update 2018. Gastroenterology 156:254, 2019. [PubMed: 30315778]

Riddle MS et al: ACG Clinical Guideline: Diagnosis, treatment, and prevention of acute diarrheal infections in adults. Am J Gastroenterol 111:602, 2016. [PubMed: 27068718]

Rubio-Tapia A et al: American College of Gastroenterology. ACG clinical guidelines: Diagnosis and management of celiac disease. Am J Gastroenterol 108:656, 2013. [PubMed: 23609613]

Smalley W et al: AGA Clinical Practice Guidelines on the laboratory evaluation of functional diarrhea and diarrhea-predominant irritable bowel syndrome in adults (IBS-D). Gastroenterology 157:851, 2019. [PubMed: 31302098]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Smalley W et al: AGA Clinical Practice Guidelines on the laboratory evaluation of functional diarrhea and diarrhea-predominant irritable bowel syndrome in adults (IBS-D). Gastroenterology 157:851, 2019. [PubMed: 31302098]

Uzzan M et al: Gastrointestinal disorders associated with common variable immune deficiency (CVID) and chronic granulomatous disease (CGD). Curr Gastroenterol Rep 18:17, 2016. [PubMed: 26951230]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 46: Diarrhea and Constipation, Michael Camilleri; Joseph A. Murray
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 47: Unintentional Weight Loss

J. Larry Jameson

INTRODUCTION

Involuntary or unintentional weight loss (UWL) is frequently insidious and can have important implications, often serving as a harbinger of serious underlying disease. Clinically important weight loss is defined as the loss of 10 pounds (4.5 kg) or >5% of one's body weight over a period of 6–12 months. UWL is encountered in up to 8% of all adult outpatients and 27% of frail persons aged ≥65 years. There is no identifiable cause in up to one-quarter of patients despite extensive investigation. Conversely, up to half of people who claim to have lost weight have no documented evidence of weight loss. People with no known cause of weight loss generally have a better prognosis than do those with known causes, particularly when the source is neoplastic. Weight loss in older persons is associated with a variety of deleterious effects, including falls and fractures, pressure ulcers, impaired immune function, and decreased functional status. Not surprisingly, significant weight loss is associated with increased mortality, which can range from 9% to as high as 38% within 1–2.5 years in the absence of clinical awareness and attention.

PHYSIOLOGY OF WEIGHT REGULATION WITH AGING

(See also Chaps. 401 and 476) Among healthy aging people, total body weight peaks in the sixth decade of life and generally remains stable until the ninth decade, after which it gradually falls. In contrast, lean body mass (fat-free mass) begins to decline at a rate of 0.3 kg per year in the third decade, and the rate of decline increases further beginning at age 60 in men and age 65 in women. These changes in lean body mass largely reflect the age-dependent decline in growth hormone secretion and, consequently, circulating levels of insulin-like growth factor type I (IGF-I) that occur with normal aging. Loss of sex steroids, at menopause in women and more gradually in men, also contributes to these changes in body composition. In the healthy elderly, an increase in fat tissue balances the loss in lean body mass until very old age, when loss of both fat and skeletal muscle occurs. Age-dependent changes also occur at the cellular level. Telomeres shorten, and body cell mass—the fat-free portion of cells—declines steadily with aging.

Between ages 20 and 80, mean energy intake is reduced by up to 1200 kcal/d in men and 800 kcal/d in women. Decreased hunger is a reflection of reduced physical activity and loss of lean body mass, producing lower demand for calories and food intake. Several important age-associated physiologic changes also predispose elderly persons to weight loss, such as declining chemosensory function (smell and taste), reduced efficiency of chewing, slowed gastric emptying, and alterations in the neuroendocrine axis, including changes in levels of leptin, cholecystokinin, neuropeptide Y, and other hormones and peptides. These changes are associated with early satiety and a decline in both appetite and the hedonistic appreciation of food. Collectively, they contribute to the "anorexia of aging." As noted below, these physiologic changes with aging may be accompanied by social isolation, poverty, and immobility, further contributing to undernutrition.

CAUSES OF UNINTENTIONAL WEIGHT LOSS

Most causes of UWL belong to one of four categories: (1) malignant neoplasms, (2) chronic inflammatory or infectious diseases, (3) metabolic disorders (e.g., hyperthyroidism and diabetes), or (4) psychiatric disorders **(Table 47-1)**. Not infrequently, more than one of these causes can be responsible for UWL. Depending upon patient populations, UWL is caused by malignant disease in a quarter of patients and by organic disease in one-third, with the remainder due to psychiatric disease, medications, or uncertain causes. Risk factors for undiagnosed cancer include a history of smoking, particularly for men, localizing symptoms, and abnormal laboratory tests.

TABLE 47-1

Causes of Involuntary Weight Loss

Cancer	Medications
Colon	Sedatives
Lung	Serotonin reuptake inhibitors
Breast	Metformin

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 47: Unintentional Weight Loss, J. Larry Jameson
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

smoking, particularly for men, localizing symptoms, and abnormal laboratory tests.

TABLE 47-1

Causes of Involuntary Weight Loss

Cancer	**Medications**
Colon	Sedatives
Hepatobiliary	Antibiotics
Hematologic	Nonsteroidal anti-inflammatory drugs
Lung	Serotonin reuptake inhibitors
Breast	Metformin
Genitourinary	Levodopa
Ovarian	Angiotensin-converting enzyme inhibitors
Prostate	Other drugs
Gastrointestinal disorders	**Disorders of the mouth and teeth**
Difficulty swallowing	Caries
Malabsorption	Dysgeusia
Peptic ulcer	**Age-related factors**
Inflammatory bowel disease	Physiologic changes
Pancreatitis	Visual impairment
Obstruction/constipation	Decreased taste and smell
Pernicious anemia	Functional disabilities
Endocrine and metabolic	**Neurologic**
Hyperthyroidism	Stroke
Diabetes mellitus	Parkinson's disease
Pheochromocytoma	Neuromuscular disorders
Adrenal insufficiency	Dementia
Cardiac disorders	**Social**
Chronic ischemia	Isolation
Chronic congestive heart failure	Poverty
Respiratory disorders	**Psychiatric and behavioral**
Emphysema	Depression
Chronic obstructive pulmonary disease	Anxiety
Renal insufficiency	Paranoia
Rheumatologic disease	Bereavement
Infections	Alcoholism
HIV	Eating disorders
Tuberculosis	Increased activity or exercise
Parasitic infection	**Idiopathic**
Subacute bacterial endocarditis	

The most common malignant causes of UWL are gastrointestinal, hepatobiliary, hematologic, lung, breast, genitourinary, ovarian, and prostate. Half of all patients with cancer lose some body weight; one-third lose more than 5% of their original body weight, and up to 20% of all cancer deaths are caused directly by cachexia (through immobility and/or cardiac/respiratory failure). The greatest incidence of weight loss is seen among patients with solid tumors. Malignancy that reveals itself through significant weight loss usually has a very poor prognosis.

In addition to malignancies, gastrointestinal diseases are among the most prominent causes of UWL. Peptic ulcer disease, inflammatory bowel disease, dysmotility syndromes, chronic pancreatitis, celiac disease, constipation, and atrophic gastritis are some of the more common entities. Oral and dental problems are easily overlooked and may manifest with halitosis, poor oral hygiene, xerostomia, inability to chew, reduced masticatory force, nonocclusion, temporomandibular joint syndrome, edentulousness, and pain due to caries or abscesses.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 47: Unintentional Weight Loss, J. Larry Jameson
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

anorexia, and vomiting. Connective tissue diseases may increase metabolic demand and disrupt nutritional balance. As the incidence of diabetes mellitus increases with aging, the associated glucosuria can contribute to weight loss. Hyperthyroidism in the elderly may have less prominent

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

nonocclusion, temporomandibular joint syndrome, edentulousness, and pain due to caries or abscesses.

Tuberculosis, fungal diseases, parasites, subacute bacterial endocarditis, and HIV are well-documented causes of UWL. Cardiovascular and pulmonary diseases cause UWL through increased metabolic demand and decreased appetite and caloric intake. Repeated surgeries may lead to weight loss because of reduced caloric intake and increased metabolic demands resulting from a systemic inflammatory response. Uremia produces nausea, anorexia, and vomiting. Connective tissue diseases may increase metabolic demand and disrupt nutritional balance. As the incidence of diabetes mellitus increases with aging, the associated glucosuria can contribute to weight loss. Hyperthyroidism in the elderly may have less prominent sympathomimetic features and may present as "apathetic hyperthyroidism" or T_3 toxicosis **(Chap. 382)**.

Neurologic injuries such as stroke, quadriplegia, and multiple sclerosis may lead to visceral and autonomic dysfunction that can impair caloric intake. Dysphagia from these neurologic insults is a common mechanism. Functional disability that compromises activities of daily living (ADLs) is a common cause of undernutrition in the elderly. Visual impairment from ophthalmic or central nervous system disorders such as a tremor can limit the ability of people to prepare and eat meals. UWL may be one of the earliest manifestations of Alzheimer's dementia.

Isolation and depression are significant causes of UWL that may manifest as an inability to care for oneself, including nutritional needs. A cytokine-mediated inflammatory metabolic cascade can be both a cause of and a manifestation of depression. Bereavement can be a cause of UWL and, when present, is often more pronounced in men. More intense forms of mental illness such as paranoid disorders may lead to delusions about food and cause weight loss. Alcoholism can be a significant source of weight loss and malnutrition.

Elderly persons living in poverty may have to choose whether to purchase food or use the money for other expenses, including medications. Screening questions can probe whether patients have run out of food or whether they routinely purchase less than they need. Institutionalization is an independent risk factor, as up to 30–50% of nursing home patients have inadequate food intake.

Medications can cause anorexia, nausea, vomiting, gastrointestinal distress, diarrhea, dry mouth, and changes in taste. This is particularly an issue in the elderly, many of whom take five or more medications.

ASSESSMENT

The four major manifestations of UWL are (1) anorexia (loss of appetite), (2) sarcopenia (loss of muscle mass), (3) cachexia (a syndrome that combines weight loss, loss of muscle and adipose tissue, anorexia, and weakness), and (4) dehydration. The current obesity epidemic adds complexity, as excess adipose tissue can mask the development of sarcopenia and delay awareness of the development of cachexia. If it is not possible to measure weight directly, a change in clothing size, corroboration of weight loss by a relative or friend, and a numeric estimate of weight loss provided by the patient are suggestive of true weight loss.

Initial assessment includes a comprehensive history and physical, a complete blood count, tests of liver enzyme levels, C-reactive protein, erythrocyte sedimentation rate, renal function studies, thyroid function tests, chest radiography, and an abdominal ultrasound **(Table 47-2)**. Age-, sex-, and risk factor–specific cancer screening tests, such as mammography and colonoscopy, should be performed **(Chap. 70)**. Patients at risk should have HIV testing. All elderly patients with weight loss should undergo screening for dementia and depression by using instruments such as the Mini-Mental State Examination and the Geriatric Depression Scale, respectively **(Chap. 477)**. The Mini Nutritional Assessment (*www.mna-elderly.com*) and the Nutrition Screening Initiative (*http://www-ncbi-nlm-nih-gov.libproxy.umflint.edu/pmc/articles/PMC1694757/*) are also available for the nutritional assessment of elderly patients. Almost all patients with a malignancy and >90% of those with other organic diseases have at least one laboratory abnormality. In patients presenting with substantial UWL, major organic and malignant diseases are unlikely when a baseline evaluation is completely normal. Careful follow-up rather than undirected testing is advised because the prognosis of weight loss of undetermined cause is generally favorable.

TABLE 47-2

Assessment and Testing for Involuntary Weight Loss

Indications	Laboratory
5% weight loss in 30 d	Complete blood count
10% weight loss in 180 d	Comprehensive electrolyte and metabolic panel, including liver and renal function tests
25% of food left uneaten after 7 d	Erythrocyte sedimentation rate

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 47: Unintentional Weight Loss, J. Larry Jameson
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

follow-up rather than undirected testing is advised because the prognosis of weight loss of undetermined cause is generally favorable.

TABLE 47-2

Assessment and Testing for Involuntary Weight Loss

Indications	**Laboratory**
5% weight loss in 30 d	Complete blood count
10% weight loss in 180 d	Comprehensive electrolyte and metabolic panel, including liver and renal function tests
Body mass index <21	Thyroid function tests
25% of food left uneaten after 7 d	Erythrocyte sedimentation rate
Change in fit of clothing	C-reactive protein
Change in appetite, smell, or taste	Ferritin
Abdominal pain, nausea, vomiting, diarrhea, constipation, dysphagia	HIV testing, if indicated
Assessment	**Radiology**
Complete physical examination, including dental evaluation	Chest x-ray Abdominal ultrasound
Medication review	
Recommended cancer screening	
Mini-Mental State Examination[a]	
Mini-Nutritional Assessment[a]	
Nutrition Screening Initiative[a]	
Simplified Nutritional Assessment Questionnaire[a]	
Observation of eating[a]	
Activities of daily living[a]	
Instrumental activities of daily living[a]	

[a]May be more specific to assess weight loss in the elderly.

TREATMENT

Unintentional Weight Loss

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 47: Unintentional Weight Loss, J. Larry Jameson
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

withdrawn or changed, if possible. For those with unexplained UWL, oral nutritional supplements such as high-energy drinks sometimes reverse weight loss. Advising patients to consume supplements between meals rather than with a meal may help minimize appetite suppression and

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Unintentional Weight Loss

The first priority in managing weight loss is to identify and treat the underlying causes. Treatment of underlying metabolic, psychiatric, infectious, or other systemic disorders may be sufficient to restore weight and functional status gradually. Medications that cause nausea or anorexia should be withdrawn or changed, if possible. For those with unexplained UWL, oral nutritional supplements such as high-energy drinks sometimes reverse weight loss. Advising patients to consume supplements between meals rather than with a meal may help minimize appetite suppression and facilitate increased overall intake. Orexigenic, anabolic, and anticytokine agents are under investigation. In selected patients, the antidepressant mirtazapine results in a significant increase in body weight, body fat mass, and leptin concentration. Patients with wasting conditions who can comply with an appropriate exercise program gain muscle protein mass, strength, and endurance and may be more capable of performing ADLs.

ACKNOWLEDGMENT

The author is grateful to Russell G. Robertson, MD, for contributions to this chapter in prior editions.

FURTHER READING

Alibhai SM et al: An approach to the management of unintentional weight loss in elderly people. CMAJ 172:773, 2005. [PubMed: 15767612]

Gaddey HL, Holder K: Unintentional weight loss in older adults. Am Fam Physician 89:718, 2014. [PubMed: 24784334]

McMinn J et al: Investigation and management of unintentional weight loss in older adults. BMJ 342:d1732, 2011. [PubMed: 21447571]

Nicholson BD et al: Prioritising primary care patients with unexpected weight loss for cancer investigation. BMJ 370:m2651, 2020. [PubMed: 32816714]

Vanderschueren S et al: The diagnostic spectrum of unintentional weight loss. Eur J Intern Med 16:160, 2005. [PubMed: 15967329]

Wong CJ: Involuntary weight loss. Med Clin North Am 98:625, 2014. [PubMed: 24758965]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 47: Unintentional Weight Loss, J. Larry Jameson
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 48: Gastrointestinal Bleeding

Loren Laine

INTRODUCTION

Gastrointestinal bleeding (GIB) presents as either overt or occult bleeding. *Overt GIB* is manifested by *hematemesis*, vomitus of red blood or "coffee-grounds" material; *melena*, black, tarry stool; and/or *hematochezia*, passage of red or maroon blood from the rectum. In the absence of overt bleeding, *occult GIB* may present with *symptoms of blood loss or anemia* such as lightheadedness, syncope, angina, or dyspnea; with iron-deficiency anemia; or a positive fecal occult blood test on colorectal cancer screening. GIB is also categorized by the site of bleeding as upper, from the esophagus, stomach, or duodenum; lower, from the colon; small intestinal; or obscure GIB if the source is unclear.

GIB is the most common gastrointestinal condition leading to hospitalization in the United States, accounting for ~513,000 admissions and $5 billion in direct costs annually. The case fatality of patients hospitalized with GIB is ~2% in the United States. Patients generally die from decompensation of other underlying illnesses rather than exsanguination.

SOURCES OF GASTROINTESTINAL BLEEDING

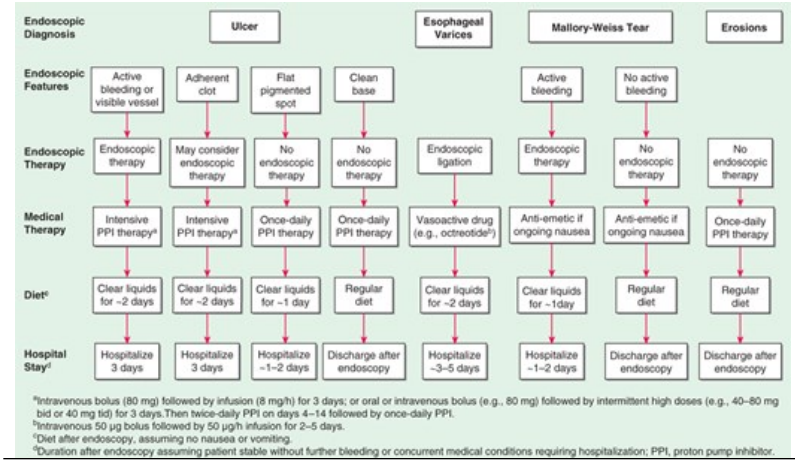
Upper Gastrointestinal Sources of Bleeding

PEPTIC ULCERS

Peptic ulcers are the most common cause of upper GIB (UGIB), accounting for ~50% of UGIB hospitalizations. Features of an ulcer at endoscopy provide important prognostic information that guides subsequent management decisions **(Fig. 48-1)**. Approximately 20% of patients with bleeding ulcers have the highest-risk findings of active bleeding or a nonbleeding visible vessel; one-third of such patients have further bleeding that requires urgent surgery if they are treated conservatively. These patients benefit from endoscopic therapy such as bipolar electrocoagulation, heater probe, injection therapy (e.g., absolute alcohol, 1:10,000 epinephrine), and/or clips with reductions in bleeding, hospital stay, mortality, and costs. In contrast, patients with clean-based ulcers have rates of serious recurrent bleeding approaching zero. If stable with no other reason for hospitalization, such patients may be discharged home after endoscopy.

FIGURE 48-1

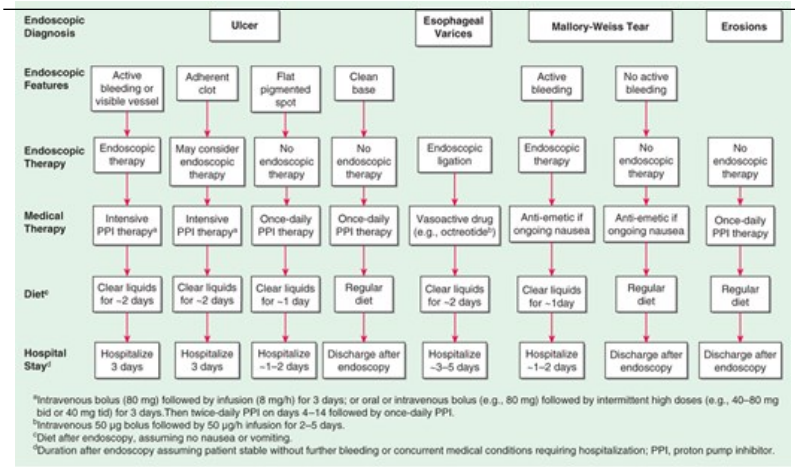
Suggested algorithm for patients with acute upper gastrointestinal bleeding based on endoscopic findings.



Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 48: Gastrointestinal Bleeding, Loren Laine
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

nonbleeding visible vessel, adherent clot) when given after endoscopic therapy. Meta-analysis of randomized trials indicates that high-dose intermittent PPIs are noninferior to constant-infusion PPI therapy and thus may be substituted. Patients with lower-risk findings (flat pigmented spot

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS



[a]Intravenous bolus (80 mg) followed by infusion (8 mg/h) for 3 days; or oral or intravenous bolus (e.g., 80 mg) followed by intermittent high doses (e.g., 40–80 mg bid or 40 mg tid) for 3 days.Then twice-daily PPI on days 4–14 followed by once-daily PPI.
[b]Intravenous 50 μg bolus followed by 50 μg/h infusion for 2–5 days.
[c]Diet after endoscopy, assuming no nausea or vomiting.
[d]Duration after endoscopy assuming patient stable without further bleeding or concurrent medical conditions requiring hospitalization; PPI, proton pump inhibitor.

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Randomized controlled trials document that high-dose, constant-infusion IV proton pump inhibitor (PPI) (80-mg bolus and 8-mg/h infusion), designed to sustain intragastric pH >6 and enhance clot stability, decreases further bleeding and mortality in patients with high-risk ulcers (active bleeding, nonbleeding visible vessel, adherent clot) when given after endoscopic therapy. Meta-analysis of randomized trials indicates that high-dose intermittent PPIs are noninferior to constant-infusion PPI therapy and thus may be substituted. Patients with lower-risk findings (flat pigmented spot or clean base) do not require endoscopic therapy and receive standard doses of oral PPI.

Approximately 10–50% of patients with bleeding ulcers rebleed within the next year if no preventive strategies are employed. Prevention of recurrent bleeding focuses on the three main factors in ulcer pathogenesis, *Helicobacter pylori*, nonsteroidal anti-inflammatory drugs (NSAIDs), and acid. Eradication of *H. pylori* in patients with bleeding ulcers decreases rebleeding rates to <5%. If a bleeding ulcer develops in a patient taking NSAIDs, the NSAIDs should be discontinued. If NSAIDs must be given, a cyclooxygenase (COX)-2 selective NSAID plus a PPI is recommended, based on results of a randomized trial. Patients with established cardiovascular disease who develop bleeding ulcers while taking low-dose aspirin for secondary prevention should restart aspirin as soon as possible after their bleeding episode (1–7 days). A randomized trial showed that immediate reinstitution of aspirin was associated with a lower 8-week mortality compared to not restarting aspirin (1% vs 13%; hazard ratio, 0.2; 95% CI, 0.1–0.6). In contrast, aspirin probably should be discontinued in most patients taking aspirin for primary prevention of cardiovascular events who develop UGIB. Patients with bleeding ulcers unrelated to *H. pylori* or NSAIDs should remain on PPI therapy indefinitely given a 42% incidence of rebleeding at 7 years without protective therapy. **Peptic ulcers are discussed in Chap. 324.**

MALLORY-WEISS TEARS

Mallory-Weiss tears account for ~2–10% of UGIB hospitalizations. The classic history is vomiting, retching, or coughing preceding hematemesis, especially in an alcoholic patient. Bleeding from these tears, which are usually on the gastric side of the gastroesophageal junction, stops spontaneously in ~80–90% of patients and recurs in only 0–10%. Endoscopic therapy is indicated for actively bleeding Mallory-Weiss tears. **Mallory-Weiss tears are discussed in Chap. 323.**

ESOPHAGEAL VARICES

The proportion of UGIB hospitalizations due to varices varies widely, from ~2–40%, depending on the population. Patients with variceal hemorrhage have poorer outcomes than patients with other sources of UGIB. Esophageal varices are treated with endoscopic ligation and an IV vasoactive medication (octreotide, somatostatin, vapreotide, terlipressin) for 2–5 days. Combination of endoscopic and medical therapy is superior to either therapy alone in decreasing rebleeding. Over the long term, treatment with nonselective beta blockers plus endoscopic ligation is recommended because the combination is more effective than either alone in reduction of recurrent esophageal variceal bleeding. Transjugular intrahepatic portosystemic shunt (TIPS) is recommended in patients who have persistent or recurrent bleeding despite endoscopic and medical therapy. TIPS also should be considered in the first 1–2 days of hospitalization for acute variceal bleeding in patients with advanced liver disease (Child-Pugh class B, Child-Pugh class C with score 10–13), because randomized trials show significant decreases in rebleeding and mortality compared with standard endoscopic and medical therapy.

Portal hypertension is also responsible for bleeding from gastric varices, varices in the small and large intestine, and portal hypertensive gastropathy

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 48: Gastrointestinal Bleeding, Loren Laine
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Erosions are endoscopically visualized breaks that are confined to the mucosa and do not cause major bleeding because arteries and veins are not

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

Portal hypertension is also responsible for bleeding from gastric varices, varices in the small and large intestine, and portal hypertensive gastropathy and enterocolopathy. Bleeding gastric varices are treated with endoscopic injection of tissue adhesive (e.g., *n*-butyl cyanoacrylate), if available; if not, TIPS is performed.

EROSIVE DISEASE

Erosions are endoscopically visualized breaks that are confined to the mucosa and do not cause major bleeding because arteries and veins are not present in the mucosa. Erosions in the esophagus, stomach, or duodenum commonly cause mild UGIB, with erosive gastritis and duodenitis accounting for perhaps ~10–15% and erosive esophagitis (primarily due to gastroesophageal reflux disease) accounting for ~1–10% of UGIB hospitalizations. The most important cause of gastric and duodenal erosions is NSAID use: ~50% of patients who chronically ingest NSAIDs may have gastric erosions. Other potential causes of gastric erosions include alcohol intake, *H. pylori* infection, and stress-related mucosal injury.

Stress-related gastric mucosal injury occurs only in extremely sick patients, such as those with serious trauma, major surgery, burns covering more than one-third of the body surface area, major intracranial disease, or severe medical illness (e.g., ventilator dependence, coagulopathy). Severe bleeding should not develop unless ulceration occurs. The mortality rate in these patients is high because of their serious underlying illnesses.

The incidence of bleeding from stress-related gastric mucosal injury has decreased dramatically in recent years, most likely due to better care of critically ill patients. A recent double-blind placebo-controlled randomized trial in 3282 intensive care patients with risk factors for GIB showed a small benefit of PPI in clinically important bleeding (2.5% vs 4.2%) without a difference in mortality or infections (e.g., *Clostridium difficile*, pneumonia). Thus, pharmacologic prophylaxis for bleeding has limited benefit but may be considered in the high-risk patients mentioned above. Meta-analyses of randomized trials suggest PPIs are more effective than H_2-receptor antagonists in reduction of overt and clinically important UGIB without differences in mortality or nosocomial pneumonia.

OTHER CAUSES

Less common causes of UGIB include neoplasms, vascular ectasias (including hereditary hemorrhagic telangiectasias [Osler-Weber-Rendu] and gastric antral vascular ectasia ["watermelon stomach"]), Dieulafoy's lesion (in which an aberrant vessel in the mucosa bleeds from a pinpoint mucosal defect), prolapse gastropathy (prolapse of proximal stomach into esophagus with retching, especially in alcoholics), aortoenteric fistulas, and hemobilia or hemosuccus pancreaticus (bleeding from the bile duct or pancreatic duct).

Small-Intestinal Sources of Bleeding

Patients without a source of GIB identified on upper endoscopy and colonoscopy were previously labeled as having obscure GIB. With the advent of improved diagnostic modalities, ~75% of GIB previously labeled obscure is now estimated to originate in the small intestine beyond the extent of a standard upper endoscopic exam. Small-intestinal GIB may account for ~5% of GIB cases. The most common causes in adults include vascular ectasias, neoplasm (e.g., gastrointestinal stromal tumor, carcinoid, adenocarcinoma, lymphoma, metastases), and NSAID-induced erosions and ulcers. Meckel's diverticulum is the most common cause of significant small-intestinal GIB in children, decreasing in frequency as a cause of bleeding with age. Other less common causes of small-intestinal GIB include Crohn's disease, infection, ischemia, vasculitis, small-bowel varices, diverticula, intussusception, Dieulafoy's lesions, aortoenteric fistulas, and duplication cysts.

Small-intestinal vascular ectasias are treated with endoscopic therapy, if possible, based on observational studies suggesting initial efficacy. However, rebleeding is common: 45% over a mean follow-up of 26 months in a systematic review. Estrogen/progesterone compounds are not recommended because a multicenter double-blind trial found no benefit in prevention of recurrent bleeding. Octreotide is used, based on positive results from case series but no randomized trials. A randomized trial reported significant benefit of thalidomide and awaits further confirmation. Other isolated lesions, such as tumors, generally require surgical resection.

Colonic Sources of Bleeding

Hemorrhoids are probably the most common cause of lower GIB (LGIB); anal fissures also cause minor bleeding and pain. If these local anal processes, which rarely require hospitalization, are excluded, the most common cause of LGIB in adults is diverticulosis. Other causes include vascular ectasias (especially in the proximal colon of patients >70 years), neoplasms (primarily adenocarcinoma), colitis (ischemic, infectious, Crohn's or ulcerative colitis, NSAID-induced colitis or ulcers), postpolypectomy bleeding, and radiation proctopathy. Rarer causes include solitary rectal ulcer syndrome, varices (most commonly rectal), lymphoid nodular hyperplasia, vasculitis, trauma, and aortocolic fistulas. In children and adolescents, the most

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 48: Gastrointestinal Bleeding, Loren Laine
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

characteristic. Case series from the United States and Europe suggest colonic diverticula stop bleeding spontaneously in ~80% of patients, with rebleeding on long-term follow-up as low as ~15% over 4-5 years. Rebleeding is substantially higher in reports from Asia. Case series suggest

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

varices (most commonly rectal), lymphoid nodular hyperplasia, vasculitis, trauma, and aortocolic fistulas. In children and adolescents, the most common colonic causes of significant GIB are inflammatory bowel disease and juvenile polyps.

Diverticular bleeding is abrupt in onset, usually painless, sometimes massive, and often from the right colon; chronic or occult bleeding is not characteristic. Case series from the United States and Europe suggest colonic diverticula stop bleeding spontaneously in ≥90% of patients, with rebleeding on long-term follow-up as low as ~15% over 4-5 years. Rebleeding is substantially higher in reports from Asia. Case series suggest endoscopic therapy may decrease recurrent bleeding in the uncommon case when colonoscopy identifies the specific bleeding diverticulum. When diverticular bleeding is found at angiography, transcatheter arterial embolization by superselective technique stops bleeding in a majority of patients. Segmental surgical resection is recommended for persistent or refractory diverticular bleeding.

Bleeding from colonic vascular ectasias may be overt or occult; it tends to be chronic and only occasionally hemodynamically significant. Endoscopic hemostatic therapy may be used in the treatment of vascular ectasias, as well as discrete bleeding ulcers and postpolypectomy bleeding. Transcatheter arterial embolization also may be attempted for persistent bleeding from vascular ectasias and other discrete lesions. Surgical therapy is generally required for major persistent or recurrent bleeding from colonic sources that cannot be treated medically, endoscopically, or angiographically. Patients with Heyde's syndrome (bleeding vascular ectasias and aortic stenosis) appear to benefit from aortic valve replacement.

APPROACH TO THE PATIENT

Gastrointestinal Bleeding

INITIAL ASSESSMENT

Measurement of the heart rate and blood pressure is the best way to initially assess a patient with GIB. Clinically significant bleeding leads to postural changes in heart rate or blood pressure, tachycardia, and, finally, recumbent hypotension. In contrast, hemoglobin does not fall immediately with acute GIB, due to proportionate reductions in plasma and red cell volumes ("people bleed whole blood"). Thus, hemoglobin may be normal or only minimally decreased at initial presentation of a severe bleeding episode. As extravascular fluid enters the vascular space to restore volume, the hemoglobin falls, but this process may take up to 72 h. Transfusion is recommended when the hemoglobin drops below 7 g/dL, based on a large randomized trial showing this restrictive transfusion strategy decreases rebleeding and death in acute UGIB compared with a transfusion threshold of 9 g/dL. Patients with slow, chronic GIB may have very low hemoglobin values despite normal blood pressure and heart rate. With the development of iron-deficiency anemia, the mean corpuscular volume is low and red blood cell distribution width is increased.

DIFFERENTIATION OF UGIB FROM LGIB

Hematemesis indicates an UGIB source. Melena indicates blood has been present in the gastrointestinal (GI) tract for ≥14 h and as long as 3–5 days. The more proximal the bleeding site, the more likely melena will occur. Hematochezia usually represents a lower GI source of bleeding, although an upper GI lesion may bleed so briskly that blood transits the bowel before melena develops. When hematochezia is the presenting symptom of UGIB, it is associated with hemodynamic instability and dropping hemoglobin. Bleeding lesions of the small bowel may present as melena or hematochezia. Other clues to UGIB include hyperactive bowel sounds and an elevated blood urea nitrogen (due to volume depletion and blood proteins absorbed in the small intestine).

A nonbloody nasogastric aspirate may be seen in ~15% of patients with UGIB who present with clinically serious hematochezia. A bile-stained appearance does not exclude UGIB because reports of bile in the aspirate are incorrect in ~50% of cases. Testing of aspirates that are not grossly bloody for occult blood is not useful.

EVALUATION AND MANAGEMENT OF UGIB

(FIG. 48-1)

Initial Risk Assessment

Baseline characteristics predictive of rebleeding and death include hemodynamic compromise (tachycardia or hypotension), increasing age, and comorbidities. Risk assessment tools may be used to identify patients with very low risk. Discharge from the emergency room with outpatient management has been suggested for patients with a Glasgow-Blatchford score (possible range 0–23, **Table 48-1**) of 0–1 because only ~1% of patients who require transfusion, require hemostatic intervention, or die have a score of 0–1.

Pre-endoscopic Medications

PPI infusion may be considered at presentation; it decreases high-risk ulcer stigmata (e.g., active bleeding) and need for endoscopic therapy but

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 48: Gastrointestinal Bleeding, Loren Laine
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

presentation. Antibiotics decrease bacterial infections, rebleeding, and mortality, and vasoactive medications may improve control of bleeding in the 12 h after presentation.

Pre-endoscopic Medications

PPI infusion may be considered at presentation; it decreases high-risk ulcer stigmata (e.g., active bleeding) and need for endoscopic therapy but does not improve clinical outcomes such as further bleeding, surgery, or death. The promotility agent erythromycin, 250 mg intravenously ~30–90 min before endoscopy, is suggested to improve visualization at endoscopy, thereby reducing the need for repeat endoscopy and hospital stay. Cirrhotic patients presenting with UGIB should be given an antibiotic (e.g., ceftriaxone) and IV vasoactive medication (e.g., octreotide) upon presentation. Antibiotics decrease bacterial infections, rebleeding, and mortality, and vasoactive medications may improve control of bleeding in the 12 h after presentation.

Endoscopy

Upper endoscopy should be performed within 24 h in most patients hospitalized with UGIB whether they have clinical features predicting low risk or high risk of further bleeding and death. Even in high-risk patients, more urgent endoscopy (performed within 6 h of gastroenterology consultation) does not improve clinical outcomes. Early endoscopy in low-risk patients (e.g., hemodynamically stable without severe comorbidities) identifies low-risk findings (e.g., clean-based ulcers, erosions, nonbleeding Mallory-Weiss tears) that allow discharge in ≥40% of patients, thereby reducing hospital stay and costs. Patients with high-risk endoscopic findings (e.g., varices, ulcers with active bleeding or a visible vessel) benefit from hemostatic therapy at endoscopy.

EVALUATION AND MANAGEMENT OF LGIB

Patients with hematochezia and hemodynamic instability should have upper endoscopy to rule out an upper GI source before evaluation of the lower GI tract **(FIG. 48-2)**.

Colonoscopy after an oral lavage solution is the procedure of choice in most patients admitted with LGIB unless bleeding is too massive, in which case angiography is recommended. Computed tomography (CT) angiography is often suggested prior to angiography to document evidence and location of active bleeding. Sigmoidoscopy is used primarily in patients <40 years old with minor bleeding. In patients with no source identified on colonoscopy, imaging studies may be employed. ^{99m}Tc-labeled red cell scan allows repeated imaging for up to 24 h and may identify the general location of bleeding. However, CT angiography is increasingly used instead because it is likely superior and more readily available. In active LGIB, angiography can detect the site of bleeding (extravasation of contrast into the gut) and permits treatment with transcatheter arterial embolization.

EVALUATION AND MANAGEMENT OF SMALL-INTESTINAL OR OBSCURE GIB

In patients with massive bleeding suspected to be from the small intestine, current guidelines suggest angiography as the initial test, with CT angiography or ^{99m}Tc-labeled red cell scan prior to angiography if the patient's clinical status permits. For others, repeat upper and lower endoscopy may be considered as the initial evaluation because second-look procedures identify a source in up to ~25% of upper endoscopies and colonoscopies; a push enteroscopy, usually performed with a pediatric colonoscope to inspect the entire duodenum and proximal jejunum, may be substituted for a repeat standard upper endoscopy. If second-look procedures are negative, evaluation of the entire small intestine is performed, usually with video capsule endoscopy. A systematic review of comparative studies showed the yield of "clinically significant findings" to be greater with capsule than push enteroscopy (56% vs 26%) or small bowel barium radiography (42% vs 6%). However, capsule endoscopy does not allow full visualization of the small intestine, tissue sampling, or application of therapy.

CT enterography may be used initially instead of video capsule in patients with possible small bowel narrowing (e.g., stricture, prior surgery or radiation, Crohn's disease) and may follow a negative video capsule for suspected small-intestinal GIB, given its higher sensitivity for small-intestinal masses.

If capsule endoscopy is positive, management is dictated by the finding. If capsule endoscopy is negative, clinically stable patients may be observed and treated with iron if iron deficiency is present, while those with ongoing bleeding (e.g., need for transfusions) undergo further testing. A second capsule endoscopy may be considered because it is reported to identify a source in up to ~50% of cases. "Deep" enteroscopy (double-balloon, single-balloon, or spiral enteroscopy) is commonly the next test after capsule endoscopy for clinically important GIB documented or suspected to be from the small intestine because it allows the endoscopist to examine, obtain specimens from, and provide therapy to much or all of the small intestine. Other imaging techniques sometimes used in evaluation of obscure GIB include ^{99m}Tc-labeled red blood cell scintigraphy, CT angiography, angiography, and ^{99m}Tc-pertechnetate scintigraphy for Meckel's diverticulum (especially in young patients). If all tests are unrevealing, intraoperative endoscopy is indicated in patients with severe recurrent or persistent bleeding requiring repeated transfusions.

POSITIVE FECAL OCCULT BLOOD TEST

Fecal occult blood testing is recommended only for colorectal cancer screening, beginning at age 45–50 years in average-risk adults. A positive test

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 48: Gastrointestinal Bleeding, Loren Laine
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

FIGURE 48-2

ACCESS Medicine®

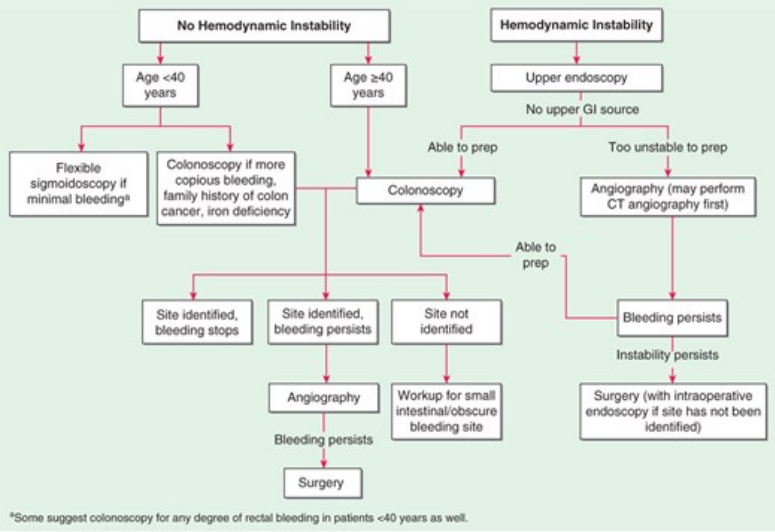
Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

POSITIVE FECAL OCCULT BLOOD TEST

Fecal occult blood testing is recommended only for colorectal cancer screening, beginning at age 45–50 years in average-risk adults. A positive test necessitates colonoscopy. If evaluation of the colon is negative, further workup is not recommended unless iron-deficiency anemia or GI symptoms are present.

FIGURE 48-2

Suggested algorithm for patients with acute lower gastrointestinal bleeding.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 48-1

Glasgow-Blatchford Score

RISK FACTORS AT ADMISSION	SCORE
Blood urea nitrogen (mg/dL)	
18.2 to <22.4	2
22.4 to <28.0	3
28.0 to <70.0	4
≥70.0	6
Hemoglobin (g/dL)	
12.0 to <13.0 (men); 10.0 to <12.0 (women)	1
10.0 to <12.0 (men)	3
<10.0	6
Systolic blood pressure (mmHg)	
100–109	1
<90	3

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 48: Gastrointestinal Bleeding, Loren Laine
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 48-1

Glasgow-Blatchford Score

RISK FACTORS AT ADMISSION	SCORE
Blood urea nitrogen (mg/dL)	
18.2 to <22.4	2
22.4 to <28.0	3
28.0 to <70.0	4
≥70.0	6
Hemoglobin (g/dL)	
12.0 to <13.0 (men); 10.0 to <12.0 (women)	1
10.0 to <12.0 (men)	3
<10.0	6
Systolic blood pressure (mmHg)	
100–109	1
90–99	2
<90	3
Heart rate (beats per minute)	
≥100	1
Melena	1
Syncope	2
Hepatic disease	2
Cardiac failure	2

FURTHER READING

Garcia-Tsao G et al: Portal hypertensive bleeding in cirrhosis: Risk stratification, diagnosis, and management: 2016 practice guidance by the American Association for the Study of Liver Diseases. Hepatology 65:310, 2017. [PubMed: 27786365]

Gurudu SR et al: The role of endoscopy in the management of suspected small-bowel bleeding. Gastrointest Endosc 85:22, 2017. [PubMed: 27374798]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 48: Gastrointestinal Bleeding, Loren Laine
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Laine L et al: ACG clinical guideline: Upper gastrointestinal and ulcer bleeding. Am J Gastroenterol 116:899, 2021. [PubMed: 33929377]

Gurudu SR et al: The role of endoscopy in the management of suspected small-bowel bleeding. Gastrointest Endosc 85:22, 2017. [PubMed: 27374798]

Krag M et al: Pantoprazole in patients at risk for gastrointestinal bleeding in the ICU. N Engl J Med 379:2199, 2018. [PubMed: 30354950]

Laine L et al: ACG clinical guideline: Upper gastrointestinal and ulcer bleeding. Am J Gastroenterol 116:899, 2021. [PubMed: 33929377]

Lau JYW et al: Timing of endoscopy for acute upper gastrointestinal bleeding. N Engl J Med 382:1299, 2020. [PubMed: 32242355]

Peery AF et al: Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: Update 2018. Gastroenterology 156:254, 2019. [PubMed: 30315778]

Stanley AJ, Laine L: Management of acute upper gastrointestinal bleeding. BMJ 364:l536, 2019. [PubMed: 30910853]

Strate LL, Gralnek KM: ACG clinical guideline: Management of patients with acute lower gastrointestinal bleeding. Am J Gastroenterol 111:459, 2016. [PubMed: 26925883]

Villaneuva C et al: Transfusion strategies for acute upper gastrointestinal bleeding. N Engl J Med 368:11, 2013. [PubMed: 23281973]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 48: Gastrointestinal Bleeding, Loren Laine
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 49: Jaundice

Savio John; Daniel S. Pratt

INTRODUCTION

Jaundice is a yellowish discoloration of body tissues resulting from the deposition of bilirubin. Tissue deposition of bilirubin occurs only in the presence of serum hyperbilirubinemia and is a sign of either liver disease or, less often, a hemolytic disorder or disorder of bilirubin metabolism. The degree of serum bilirubin elevation can be estimated by physical examination. Slight increases in serum bilirubin level are best detected by examining the sclerae for icterus. Sclerae have a particular affinity for bilirubin due to their high elastin content, and the presence of scleral icterus indicates a serum bilirubin level of at least 51 µmol/L (3 mg/dL). The ability to detect scleral icterus is made more difficult if the examining room has fluorescent lighting. If the examiner suspects scleral icterus, a second site to examine is underneath the tongue. As serum bilirubin levels rise, the skin will eventually become yellow in light-skinned patients and even green if the process is long-standing; the green color is produced by oxidation of bilirubin to biliverdin.

The differential diagnosis for yellowing of the skin is limited. In addition to jaundice, it includes carotenoderma; the use of drugs including quinacrine, sunitinib, and sorafenib; and excessive exposure to phenols. Carotenoderma, a yellow coloring of the skin, is associated with diabetes, hypothyroidism, and anorexia nervosa, but most commonly, it is caused by the ingestion of an excessive amounts of vegetables and fruits such as carrots, leafy vegetables, squash, peaches, and oranges that contain carotene. In jaundice, the yellow coloration of the skin is uniformly distributed over the body, whereas in carotenoderma, the pigment is concentrated on the palms, soles, forehead, and nasolabial folds. Carotenoderma can be distinguished from jaundice by the sparing of the sclerae. Quinacrine causes a yellow discoloration of the skin in 4–37% of patients treated with it. It has also been reported with the use of the tyrosine kinase inhibitors sunitinib and sorafenib.

Another sensitive indicator of increased serum bilirubin is darkening of the urine, which is due to the renal excretion of conjugated bilirubin. Patients often describe their urine as tea- or cola-colored. Bilirubinuria indicates an elevation of the direct serum bilirubin fraction and, therefore, the presence of liver or biliary disease.

Serum bilirubin levels increase when an imbalance exists between bilirubin production and clearance. A logical evaluation of the patient who is jaundiced requires an understanding of bilirubin production and metabolism.

PRODUCTION AND METABOLISM OF BILIRUBIN

(See Chap. 338) Bilirubin, a tetrapyrrole pigment, is a breakdown product of heme (ferroprotoporphyrin IX). About 80–85% of the 4 mg/kg body weight of bilirubin produced each day is derived from the breakdown of hemoglobin in senescent red blood cells. The remainder comes from prematurely destroyed erythroid cells in bone marrow and from the turnover of hemoproteins such as myoglobin and cytochromes found in tissues throughout the body.

The formation of bilirubin occurs in reticuloendothelial cells, primarily in the spleen and liver. The first reaction, catalyzed by the microsomal enzyme heme oxygenase, oxidatively cleaves the α bridge of the porphyrin group and opens the heme ring. The end products of this reaction are biliverdin, carbon monoxide, and iron. The second reaction, catalyzed by the cytosolic enzyme biliverdin reductase, reduces the central methylene bridge of biliverdin and converts it to bilirubin. Bilirubin formed in the reticuloendothelial cells is virtually insoluble in water due to tight internal hydrogen bonding between the water-soluble moieties of bilirubin—that is, the bonding of the propionic acid carboxyl groups of one dipyrrolic half of the molecule with the imino and lactam groups of the opposite half. This configuration blocks solvent access to the polar residues of bilirubin and places the hydrophobic residues on the outside. To be transported in blood, bilirubin must be solubilized. Solubilization is accomplished by the reversible, noncovalent binding of bilirubin to albumin. Unconjugated bilirubin bound to albumin is transported to the liver. There, the bilirubin—but not the albumin—is taken up by hepatocytes via a process that at least partly involves carrier-mediated membrane transport. No specific bilirubin transporter has yet been identified **(Chap. 338, Fig. 338-1)**.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

endoplasmic reticulum, bilirubin is made aqueous soluble by conjugation to glucuronic acid, a process that disrupts the hydrophobic internal hydrogen bonds and yields bilirubin monoglucuronide and diglucuronide. The conjugation of glucuronic acid to bilirubin is catalyzed by bilirubin

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

albumin—is taken up by hepatocytes via a process that at least partly involves carrier-mediated membrane transport. No specific bilirubin transporter has yet been identified **(Chap. 338, Fig. 338-1)**.

After entering the hepatocyte, unconjugated bilirubin is bound in the cytosol to several proteins including proteins in the glutathione-S-transferase superfamily. These proteins serve both to reduce efflux of bilirubin back into the serum and to present the bilirubin for conjugation. In the endoplasmic reticulum, bilirubin is made aqueous soluble by conjugation to glucuronic acid, a process that disrupts the hydrophobic internal hydrogen bonds and yields bilirubin monoglucuronide and diglucuronide. The conjugation of glucuronic acid to bilirubin is catalyzed by bilirubin uridine diphosphate-glucuronosyl transferase (UDPGT). The now-hydrophilic bilirubin conjugates diffuse from the endoplasmic reticulum to the canalicular membrane, where bilirubin monoglucuronide and diglucuronide are actively transported into canalicular bile by an energy-dependent mechanism involving the multidrug resistance–associated protein 2 (MRP2). A portion of bilirubin glucuronides is transported into the sinusoids and portal circulation by MRP3 and is subjected to reuptake into the hepatocyte by the sinusoidal organic anion transport protein 1B1 (OATP1B1) and OATP1B3. The conjugated bilirubin excreted into bile drains into the duodenum and passes unchanged through the proximal small bowel. Conjugated bilirubin is not reabsorbed by the intestinal mucosa due to its hydrophilicity and increased molecular size. When the conjugated bilirubin reaches the distal ileum and colon, it is hydrolyzed to unconjugated bilirubin by bacterial β-glucuronidases. The unconjugated bilirubin is reduced by normal gut bacteria to form a group of colorless tetrapyrroles called *urobilinogens* and other products, the nature and relative amounts of which depend on the bacterial flora. About 80–90% of these products are excreted in feces, either unchanged or oxidized to orange derivatives called *urobilins*. The remaining 10–20% of the urobilinogens undergo enterohepatic cycling. A small fraction (usually <3 mg/dL) escapes hepatic uptake, filters across the renal glomerulus, and is excreted in urine. Increased urinary excretion of urobilinogen can be due to increased bilirubin production, increased hepatic reabsorption of urobilinogen from the colon, or decreased hepatic clearance of urobilinogen.

MEASUREMENT OF SERUM BILIRUBIN

The terms *direct* and *indirect* bilirubin—that is, conjugated and unconjugated bilirubin, respectively—are based on the original van den Bergh reaction. This assay, or a variation of it, is still used in most clinical chemistry laboratories to determine the serum bilirubin level. In this assay, bilirubin is exposed to diazotized sulfanilic acid and splits into two relatively stable dipyrrylmethene azopigments that absorb maximally at 540 nm, allowing photometric analysis. The direct fraction is that which reacts with diazotized sulfanilic acid in the absence of an accelerator substance such as alcohol. The direct fraction provides an approximation of the conjugated bilirubin level in serum. The *total* serum bilirubin is the amount that reacts after the addition of alcohol. The indirect fraction is the difference between the total and the direct bilirubin levels and provides an estimate of the unconjugated bilirubin in serum. Unconjugated bilirubin also reacts with diazo reagents, albeit slowly, even when the accelerator is absent. Thus, the calculated indirect bilirubin may underestimate the true amount of unconjugated bilirubin in circulation.

With the van den Bergh method, the normal serum bilirubin concentration usually is between 17 and 26 µmol/L (1 and 1.5 mg/dL). Total serum bilirubin concentrations are between 3.4 and 15.4 µmol/L (0.2 and 0.9 mg/dL) in 95% of a normal population. Unconjugated hyperbilirubinemia is present when the direct fraction is <15% of the total serum bilirubin. The presence of even limited amounts of true conjugated bilirubin in serum suggests significant hepatobiliary pathology. As conjugated hyperbilirubinemia is always associated with bilirubinuria (except in the presence of delta bilirubin in prolonged cholestasis when jaundice is overt), detection of bilirubin in urine via dipstick test is extremely helpful to confirm the presence of conjugated hyperbilirubinemia in a patient with mildly elevated direct fraction.

Several new techniques, although less convenient to perform, have added considerably to our understanding of bilirubin metabolism. First, studies using these methods demonstrate that, in normal persons or those with Gilbert's syndrome, almost 100% of the serum bilirubin is unconjugated; <3% is monoconjugated bilirubin. Second, in jaundiced patients with hepatobiliary disease, the total serum bilirubin concentration measured by these new, more accurate methods is lower than the values found with diazo methods. This finding suggests that there are diazo-positive compounds distinct from bilirubin in the serum of patients with hepatobiliary disease. Third, these studies indicate that, in jaundiced patients with hepatobiliary disease, monoglucuronides of bilirubin predominate over diglucuronides. Fourth, part of the direct-reacting bilirubin fraction includes conjugated bilirubin that is covalently linked to albumin. This albumin-linked fraction of conjugated bilirubin (*delta fraction, delta bilirubin*, or *biliprotein*) represents an important fraction of total serum bilirubin in patients with cholestasis and hepatobiliary disorders. The delta bilirubin is formed in serum when hepatic excretion of bilirubin glucuronides is impaired and the glucuronides accumulate in serum. By virtue of its tight binding to albumin, the clearance rate of delta bilirubin from serum approximates the half-life of albumin (12–14 days) rather than the short half-life of bilirubin (about 4 h).

The prolonged half-life of albumin-bound conjugated bilirubin accounts for two previously unexplained enigmas in jaundiced patients with liver disease: (1) that some patients with conjugated hyperbilirubinemia do not exhibit bilirubinuria during the recovery phase of their disease because the delta bilirubin, although conjugated, is covalently bound to albumin and therefore not filtered by the renal glomeruli, and (2) that the elevated serum

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

MEASUREMENT OF URINE BILIRUBIN

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

disease: (1) that some patients with conjugated hyperbilirubinemia do not exhibit bilirubinuria during the recovery phase of their disease because the delta bilirubin, although conjugated, is covalently bound to albumin and therefore not filtered by the renal glomeruli, and (2) that the elevated serum bilirubin level declines more slowly than expected in some patients who otherwise appear to be recovering satisfactorily. Late in the recovery phase of hepatobiliary disorders, all the conjugated bilirubin may be in the albumin-linked form.

MEASUREMENT OF URINE BILIRUBIN

Unconjugated bilirubin is always bound to albumin in the serum, is not filtered by the kidney, and is not found in the urine. Conjugated bilirubin is filtered at the glomerulus, and the majority is reabsorbed by the proximal tubules; a small fraction is excreted in the urine. Any bilirubin found in the urine is conjugated bilirubin. The presence of bilirubinuria on urine dipstick test (Ictotest) indicates an elevation of the conjugated bilirubin fraction that cannot be excreted from the liver and implies the presence of hepatobiliary disease. A false-negative result is possible in patients with prolonged cholestasis due to the predominance of delta bilirubin, which is covalently bound to albumin and therefore not filtered by the renal glomeruli.

APPROACH TO THE PATIENT

Jaundice

The goal of this chapter is not to provide an encyclopedic review of every condition that causes jaundice. Rather, the chapter is intended to offer a framework that helps a physician to evaluate the patient with jaundice in a logical way **(Fig. 49-1)**.

The initial step is to perform appropriate blood tests in order to determine whether the patient has an isolated elevation of serum bilirubin. If so, is the bilirubin elevation due to an increased unconjugated or conjugated fraction? If the hyperbilirubinemia is accompanied by other liver test abnormalities, is the disorder hepatocellular or cholestatic? If cholestatic, is it intra- or extrahepatic? These questions can all be answered with a thoughtful history, physical examination, and interpretation of laboratory and radiologic tests and procedures.

The bilirubin present in serum represents a balance between input from the production of bilirubin and hepatic/biliary removal of the pigment. Hyperbilirubinemia may result from (1) overproduction of bilirubin; (2) impaired uptake, conjugation, or excretion of bilirubin; or (3) regurgitation of unconjugated or conjugated bilirubin from damaged hepatocytes or bile ducts. An increase in unconjugated bilirubin in serum results from overproduction, impaired uptake, or conjugation of bilirubin. An increase in conjugated bilirubin is due to decreased excretion into the bile ductules or backward leakage of the pigment. The initial steps in evaluating the patient with jaundice are to determine (1) whether the hyperbilirubinemia is predominantly conjugated or unconjugated in nature and (2) whether other biochemical liver tests are abnormal. The thoughtful interpretation of limited data permits a rational evaluation of the patient (Fig. 49-1). The following discussion will focus solely on the evaluation of the adult patient with jaundice.

ISOLATED ELEVATION OF SERUM BILIRUBIN

Unconjugated Hyperbilirubinemia

The differential diagnosis of isolated unconjugated hyperbilirubinemia is limited **(Table 49-1)**. The critical determination is whether the patient is suffering from a hemolytic process resulting in an overproduction of bilirubin (hemolytic disorders and ineffective erythropoiesis) or from impaired hepatic uptake/conjugation of bilirubin (drug effect or genetic disorders).

Hemolytic disorders that cause excessive heme production may be either inherited or acquired. Inherited disorders include spherocytosis, sickle cell anemia, thalassemia, and deficiency of red cell enzymes such as pyruvate kinase and glucose-6-phosphate dehydrogenase. In these conditions, the serum bilirubin level rarely exceeds 86 µmol/L (5 mg/dL). Higher levels may occur when there is coexistent renal or hepatocellular dysfunction or in acute hemolysis, such as a sickle cell crisis. In evaluating jaundice in patients with chronic hemolysis, it is important to remember the high incidence of pigmented (calcium bilirubinate) gallstones found in these patients, which increases the likelihood of choledocholithiasis as an alternative explanation for hyperbilirubinemia.

Acquired hemolytic disorders include microangiopathic hemolytic anemia (e.g., hemolytic-uremic syndrome), paroxysmal nocturnal hemoglobinuria, spur cell anemia, immune hemolysis, and parasitic infections (e.g., malaria and babesiosis). Ineffective erythropoiesis occurs in cobalamin, folate, and iron deficiencies. Resorption of hematomas and massive blood transfusions both can result in increased hemoglobin release and overproduction of bilirubin.

In the absence of hemolysis, the physician should consider a problem with the hepatic uptake or conjugation of bilirubin. Certain drugs, including rifampin and probenecid, may cause unconjugated hyperbilirubinemia by diminishing hepatic uptake of bilirubin. Impaired bilirubin conjugation

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

kernicterus, frequently leading to death in infancy or childhood. These patients have a complete absence of bilirubin UDP-GT activity, are totally unable to conjugate bilirubin; and hence cannot excrete it.

In the absence of hemolysis, the physician should consider a problem with the hepatic uptake or conjugation of bilirubin. Certain drugs, including rifampin and probenecid, may cause unconjugated hyperbilirubinemia by diminishing hepatic uptake of bilirubin. Impaired bilirubin conjugation occurs in three genetic conditions: Crigler-Najjar syndrome types I and II and Gilbert's syndrome. *Crigler-Najjar type I* is an exceptionally rare condition found in neonates and characterized by severe jaundice (bilirubin >342 µmol/L [>20 mg/dL]) and neurologic impairment due to kernicterus, frequently leading to death in infancy or childhood. These patients have a complete absence of bilirubin UDPGT activity; are totally unable to conjugate bilirubin; and hence cannot excrete it.

Crigler-Najjar type II is somewhat more common than type I. Patients live into adulthood with serum bilirubin levels of 103–428 µmol/L (6–25 mg/dL). In these patients, mutations in the bilirubin *UDPGT* gene cause the reduction—typically ≤10%—of the enzyme's activity. Bilirubin UDPGT activity can be induced by the administration of phenobarbital, which can reduce serum bilirubin levels in these patients. Despite marked jaundice, these patients usually survive into adulthood, although they may be susceptible to kernicterus under the stress of concurrent illness or surgery.

Gilbert's syndrome is also marked by the impaired conjugation of bilirubin due to reduced bilirubin UDPGT activity (typically 10–35% of normal). Patients with Gilbert's syndrome have mild unconjugated hyperbilirubinemia, with serum levels almost always <103 µmol/L (6 mg/dL). The serum levels may fluctuate, and jaundice is often identified only during periods of stress, concurrent illness, alcohol use, or fasting. Unlike both Crigler-Najjar syndromes, Gilbert's syndrome is very common. The reported incidence is 3–7% of the population, with males predominating over females by a ratio of 1.5–7:1.

Conjugated Hyperbilirubinemia

Elevated conjugated hyperbilirubinemia is found in two rare inherited conditions: *Dubin-Johnson syndrome* and *Rotor syndrome* (Table 49-1). Patients with either condition present with asymptomatic jaundice. The defect in Dubin-Johnson syndrome is the presence of mutations in the gene for MRP2. These patients have altered excretion of bilirubin into the bile ducts. Rotor syndrome may represent a deficiency of the major hepatic drug reuptake transporters OATP1B1 and OATP1B3. Differentiating between these syndromes is possible but is clinically unnecessary due to their benign nature.

ELEVATION OF SERUM BILIRUBIN WITH OTHER LIVER TEST ABNORMALITIES

The remainder of this chapter will focus on the evaluation of patients with conjugated hyperbilirubinemia in the setting of other liver test abnormalities. This group of patients can be divided into those with a primary hepatocellular process and those with intra- or extrahepatic cholestasis. This distinction, which is based on the history and physical examination as well as the pattern of liver test abnormalities, guides the clinician's evaluation (Fig. 49-1).

History

A complete medical history is perhaps the single most important part of the evaluation of the patient with unexplained jaundice. Important considerations include the use of or exposure to any chemical or medication, whether physician-prescribed, over-the-counter, complementary, or alternative medicines (e.g., herbal and vitamin preparations) or other drugs such as anabolic steroids. The patient should be carefully questioned about possible parenteral exposures, including transfusions, intravenous and intranasal drug use, tattooing, and sexual activity. Other important points include recent travel history; exposure to people with jaundice; exposure to possibly contaminated foods; occupational exposure to hepatotoxins; alcohol consumption; the duration of jaundice; and the presence of any accompanying signs and symptoms, such as arthralgias, myalgias, rash, anorexia, weight loss, abdominal pain, fever, pruritus, and changes in the urine and stool. While none of the latter manifestations is specific for any one condition, any of them can suggest a diagnosis. A history of arthralgias and myalgias predating jaundice suggests hepatitis, either viral or drug related. Jaundice associated with the sudden onset of severe right-upper-quadrant pain and shaking chills suggests choledocholithiasis and ascending cholangitis.

Physical Examination

The general assessment should include evaluation of the patient's nutritional status. Temporal and proximal muscle wasting suggests long-standing disease such as pancreatic cancer or cirrhosis. Stigmata of chronic liver disease, including spider nevi, palmar erythema, gynecomastia, caput medusae, Dupuytren's contractures, parotid gland enlargement, and testicular atrophy, are commonly seen in advanced alcohol-related cirrhosis and occasionally in other types of cirrhosis. An enlarged left supraclavicular node (Virchow's node) or a periumbilical nodule (Sister Mary Joseph's nodule) suggests an abdominal malignancy. Jugular venous distention, a sign of right-sided heart failure, suggests hepatic congestion. Right pleural effusion even in the absence of clinically apparent ascites may be seen in advanced cirrhosis.

The abdominal examination should focus on the size and consistency of the liver, on whether the spleen is palpable and hence enlarged, and on whether ascites is present. Patients with cirrhosis may have an enlarged left lobe of the liver, which is felt below the xiphoid, and an enlarged spleen. Shifting dullness with respiratory excursion on inspiration (Talney's sign), suggests ascites in the presence of jaundice suggests either cirrhosis or malignancy with peritoneal spread.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR

The abdominal examination should focus on the size and consistency of the liver, on whether the spleen is palpable and hence enlarged, and on whether ascites is present. Patients with cirrhosis may have an enlarged left lobe of the liver, which is felt below the xiphoid, and an enlarged spleen. A grossly enlarged nodular liver or an obvious abdominal mass suggests malignancy. An enlarged tender liver could signify viral or alcoholic hepatitis; an infiltrative process such as amyloidosis; or, less often, an acutely congested liver secondary to right-sided heart failure. Severe right-upper-quadrant tenderness with respiratory arrest on inspiration (Murphy's sign) suggests cholecystitis. Ascites in the presence of jaundice suggests either cirrhosis or malignancy with peritoneal spread.

Laboratory Tests

A battery of tests are helpful in the initial evaluation of a patient with unexplained jaundice. These include total and direct serum bilirubin measurement with fractionation; determination of serum aminotransferase, alkaline phosphatase, and albumin concentrations; and prothrombin time tests. Enzyme tests (alanine aminotransferase [ALT], aspartate aminotransferase [AST], and alkaline phosphatase [ALP]) are helpful in differentiating between a hepatocellular process and a cholestatic process **(Table 337-1;** Fig. 49-1)—a critical step in determining what additional workup is indicated. Patients with a hepatocellular process generally have a rise in the aminotransferases that is disproportionate to that in ALP, whereas patients with a cholestatic process have a rise in ALP that is disproportionate to that of the aminotransferases. The serum bilirubin can be prominently elevated in both hepatocellular and cholestatic conditions and therefore is not necessarily helpful in differentiating between the two.

In addition to enzyme tests, all jaundiced patients should have additional blood tests—specifically, an albumin level and a prothrombin time—to assess liver function. A low albumin level suggests a chronic process such as cirrhosis or cancer. A normal albumin level is suggestive of a more acute process such as viral hepatitis or choledocholithiasis. An elevated prothrombin time indicates either vitamin K deficiency due to prolonged jaundice and malabsorption of vitamin K or significant hepatocellular dysfunction. The failure of the prothrombin time to correct with parenteral administration of vitamin K indicates severe hepatocellular injury.

The results of the bilirubin, enzyme, albumin, and prothrombin time tests will usually indicate whether a jaundiced patient has a hepatocellular or a cholestatic disease and offer some indication of the duration and severity of the disease. The causes and evaluations of hepatocellular and cholestatic diseases are quite different.

Hepatocellular Conditions

Hepatocellular diseases that can cause jaundice include viral hepatitis, drug or environmental toxicity, alcohol, and end-stage cirrhosis from any cause **(Table 49-2)**. Wilson's disease occurs primarily in young adults. Autoimmune hepatitis is typically seen in young to middle-aged women but may affect men and women of any age. Alcoholic hepatitis can be differentiated from viral and toxin-related hepatitis by the pattern of the aminotransferases: patients with alcoholic hepatitis typically have an AST-to-ALT ratio of at least 2:1, and the AST level rarely exceeds 300 U/L. Patients with acute viral hepatitis and toxin-related injury severe enough to produce jaundice typically have aminotransferase levels >500 U/L, with the ALT greater than or equal to the AST. While ALT and AST values <8 times normal may be seen in either hepatocellular or cholestatic liver disease, values 25 times normal or higher are seen primarily in acute hepatocellular diseases. Patients with jaundice from cirrhosis can have normal or only slightly elevated aminotransferase levels.

When the clinician determines that a patient has a hepatocellular disease, appropriate testing for acute viral hepatitis includes a hepatitis A IgM antibody assay, a hepatitis B surface antigen and core IgM antibody assay, a hepatitis C viral RNA test, and, depending on the circumstances, a hepatitis E IgM antibody assay. The hepatitis C antibody can take up to 6 weeks to become detectable, making it an unreliable test if acute hepatitis C is suspected. Studies for hepatitis D, Epstein-Barr virus (EBV), and cytomegalovirus (CMV) may also be indicated. Ceruloplasmin is the initial screening test for Wilson's disease. Testing for autoimmune hepatitis usually includes antinuclear antibody and anti-smooth muscle antibody assays and measurement of specific immunoglobulins.

Drug-induced hepatocellular injury can be classified as either predictable or unpredictable. Predictable drug reactions are dose-dependent and affect all patients who ingest a toxic dose of the drug in question. The classic example is acetaminophen hepatotoxicity. Unpredictable or idiosyncratic drug reactions are not dose-dependent and occur in a minority of patients. A great number of drugs can cause idiosyncratic hepatic injury. Environmental toxins are also an important cause of hepatocellular injury. Examples include industrial chemicals such as vinyl chloride, herbal preparations containing pyrrolizidine alkaloids (Jamaica bush tea) or kava, and the mushrooms *Amanita phalloides* and *A. verna*, which contain highly hepatotoxic amatoxins.

Cholestatic Conditions

When the pattern of the liver tests suggests a cholestatic disorder, the first step is to determine whether it is intra- or extrahepatic cholestasis (Fig. 49-1). Distinguishing intrahepatic from extrahepatic cholestasis may be difficult. History, physical examination, and laboratory tests often are not

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

common bile duct or in patients with cirrhosis or primary sclerosing cholangitis (PSC), in which scarring prevents the intrahepatic ducts from dilating.

herbal preparations containing pyrrolizidine alkaloids (Jamaica bush tea) or kava, and the mushrooms *Amanita phalloides* and *A. verna*, which

When the pattern of the liver tests suggests a cholestatic disorder, the first step is to determine whether it is intra- or extrahepatic cholestasis (Fig. 49-1). Distinguishing intrahepatic from extrahepatic cholestasis may be difficult. History, physical examination, and laboratory tests often are not helpful. The next appropriate test is an ultrasound. The ultrasound is inexpensive, does not expose the patient to ionizing radiation, and can detect dilation of the intra- and extrahepatic biliary tree with a high degree of sensitivity and specificity. The absence of biliary dilation suggests intrahepatic cholestasis, while its presence indicates extrahepatic cholestasis. False-negative results occur in patients with partial obstruction of the common bile duct or in patients with cirrhosis or primary sclerosing cholangitis (PSC), in which scarring prevents the intrahepatic ducts from dilating.

Although ultrasonography may indicate extrahepatic cholestasis, it rarely identifies the site or cause of obstruction. The distal common bile duct is a particularly difficult area to visualize by ultrasound because of overlying bowel gas. Appropriate next tests include computed tomography (CT), magnetic resonance cholangiopancreatography (MRCP), endoscopic retrograde cholangiopancreatography (ERCP), percutaneous transhepatic cholangiography (PTC), and endoscopic ultrasound (EUS). CT and MRCP are better than ultrasonography for assessing the head of the pancreas and for identifying choledocholithiasis in the distal common bile duct, particularly when the ducts are not dilated. ERCP is the "gold standard" for identifying choledocholithiasis. Beyond its diagnostic capabilities, ERCP allows therapeutic interventions, including the removal of common bile duct stones and the placement of stents. PTC can provide the same information as ERCP and it also allows for intervention in patients in whom ERCP is unsuccessful due to proximal biliary obstruction or altered gastrointestinal anatomy. MRCP has replaced ERCP as the initial diagnostic test in most cases. EUS displays sensitivity and specificity comparable to that of MRCP in the detection of bile duct obstruction and allows biopsy of suspected malignant lesions.

In patients with apparent *intrahepatic cholestasis*, the diagnosis is often made by serologic testing in combination with a liver biopsy. The list of possible causes of intrahepatic cholestasis is long and varied **(Table 49-3)**. A number of conditions that typically cause a hepatocellular pattern of injury can also present as a cholestatic variant. Both hepatitis B and C viruses can cause cholestatic hepatitis (fibrosing cholestatic hepatitis). This disease variant has been reported in patients who have undergone solid organ transplantation. Hepatitis A and E, alcoholic hepatitis, and EBV or CMV infections may also present as cholestatic liver disease.

Drugs may cause intrahepatic cholestasis that is usually reversible after discontinuation of the offending agent, although it may take many months for cholestasis to resolve. Drugs most commonly associated with cholestasis are the anabolic and contraceptive steroids. Cholestatic hepatitis has been reported with chlorpromazine, imipramine, tolbutamide, sulindac, cimetidine, and erythromycin estolate. It also occurs in patients taking trimethoprim; sulfamethoxazole; and penicillin-based antibiotics such as ampicillin, dicloxacillin, and clavulanic acid. Rarely, cholestasis may be chronic and associated with progressive fibrosis despite early discontinuation of the offending drug. Chronic cholestasis has been associated with chlorpromazine and prochlorperazine.

Primary biliary cholangitis is an autoimmune disease predominantly affecting women and characterized by progressive destruction of interlobular bile ducts. The diagnosis is made by the detection of antimitochondrial antibody, which is found in 95% of patients. *Primary sclerosing cholangitis* is characterized by the destruction and fibrosis of larger bile ducts. The diagnosis of PSC is made with cholangiography (either MRCP or ERCP), which demonstrates the pathognomonic segmental strictures. Approximately 75% of patients with PSC also have inflammatory bowel disease.

The *vanishing bile duct syndrome* and *adult bile ductopenia* are rare conditions in which a decreased number of bile ducts are seen in liver biopsy specimens. This histologic picture is also seen in patients who develop chronic rejection after liver transplantation and in those who develop graft-versus-host disease after bone marrow transplantation. Vanishing bile duct syndrome also occurs in rare cases of sarcoidosis, in patients taking certain drugs (including chlorpromazine), and idiopathically.

There are also familial forms of intrahepatic cholestasis. The familial intrahepatic cholestatic syndromes include *progressive familial intrahepatic cholestasis* (PFIC) *types 1–3* and *benign recurrent intrahepatic cholestasis* (BRIC) *types 1 and 2*. BRIC is characterized by episodic attacks of pruritus, cholestasis, and jaundice beginning at any age, which can be debilitating but does not lead to chronic liver disease. Serum bile acids are elevated during episodes, but serum γ-glutamyltransferase (γ-GT) activity is normal. PFIC disorders begin at childhood and are progressive in nature. All three types of PFIC are associated with progressive cholestasis, elevated levels of serum bile acids, and similar phenotypes but different genetic mutations. Only type 3 PFIC is associated with high levels of γ-GT. *Cholestasis of pregnancy* occurs in the second and third trimesters and resolves after delivery. Its cause is unknown, but the condition is probably inherited, and cholestasis can be triggered by estrogen administration.

Other causes of intrahepatic cholestasis include total parenteral nutrition (TPN); nonhepatobiliary sepsis; benign postoperative cholestasis; and a paraneoplastic syndrome associated with a number of different malignancies, including Hodgkin's disease, medullary thyroid cancer, renal cell cancer, renal sarcoma, T-cell lymphoma, prostate cancer, and several gastrointestinal malignancies. The term *Stauffer's syndrome* has been used

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

intrahepatic and extrahepatic cholestasis. Jaundice is a late finding in heart failure caused by hepatic congestion and hepatocellular hypoxia. Ischemic hepatitis is a distinct entity of acute hypoperfusion characterized by an acute and dramatic elevation in the serum aminotransferases

paraneoplastic syndrome associated with a number of different malignancies, including Hodgkin's disease, medullary thyroid cancer, renal cell cancer, renal sarcoma, T-cell lymphoma, prostate cancer, and several gastrointestinal malignancies. The term *Stauffer's syndrome* has been used for intrahepatic cholestasis specifically associated with renal cell cancer. In patients developing cholestasis in the intensive care unit, the major considerations should be sepsis, ischemic hepatitis ("shock liver"), and TPN-related jaundice. Jaundice occurring after bone marrow transplantation is most likely due to veno-occlusive disease or graft-versus-host disease. In addition to hemolysis, sickle cell disease may cause intrahepatic and extrahepatic cholestasis. Jaundice is a late finding in heart failure caused by hepatic congestion and hepatocellular hypoxia. Ischemic hepatitis is a distinct entity of acute hypoperfusion characterized by an acute and dramatic elevation in the serum aminotransferases followed by a gradual peak in serum bilirubin.

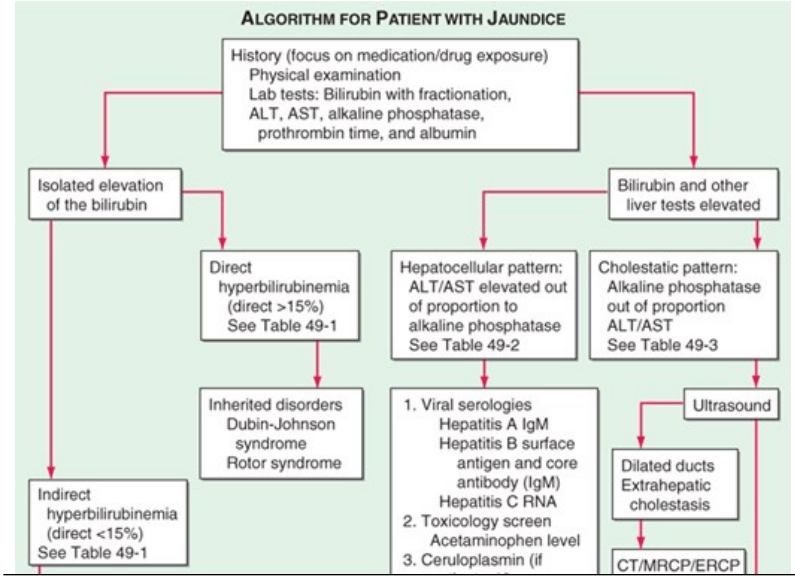
Jaundice with associated liver dysfunction can be seen in severe cases of *Plasmodium falciparum* malaria. The jaundice in these cases is due to a combination of indirect hyperbilirubinemia from hemolysis and both cholestatic and hepatocellular jaundice. Weil's disease, a severe presentation of leptospirosis, is marked by jaundice with renal failure, fever, headache, and muscle pain.

Causes of *extrahepatic cholestasis* can be split into malignant and benign (Table 49-3). Malignant causes include pancreatic, gallbladder, and ampullary cancers as well as cholangiocarcinoma. This last malignancy is most commonly associated with PSC and is exceptionally difficult to diagnose because its appearance is often identical to that of PSC. Pancreatic and gallbladder tumors as well as cholangiocarcinoma are rarely resectable and have poor prognoses. Ampullary carcinoma has the highest surgical cure rate of all the tumors that present as painless jaundice. Hilar lymphadenopathy due to metastases from other cancers may cause obstruction of the extrahepatic biliary tree.

Choledocholithiasis is the most common cause of extrahepatic cholestasis. The clinical presentation can range from mild right-upper-quadrant discomfort with only minimal elevations of enzyme test values to ascending cholangitis with jaundice, sepsis, and circulatory collapse. PSC may occur with clinically important strictures limited to the extrahepatic biliary tree. IgG4-associated cholangitis is marked by stricturing of the biliary tree. It is critical that the clinician differentiate this condition from PSC as it is responsive to glucocorticoid therapy. In rare instances, chronic pancreatitis causes strictures of the distal common bile duct, where it passes through the head of the pancreas. AIDS cholangiopathy is a condition that is usually due to infection of the bile duct epithelium with CMV or cryptosporidia and has a cholangiographic appearance similar to that of PSC. The affected patients usually present with greatly elevated serum alkaline phosphatase levels (mean, 800 IU/L), but the bilirubin level is often near normal. These patients do not typically present with jaundice.

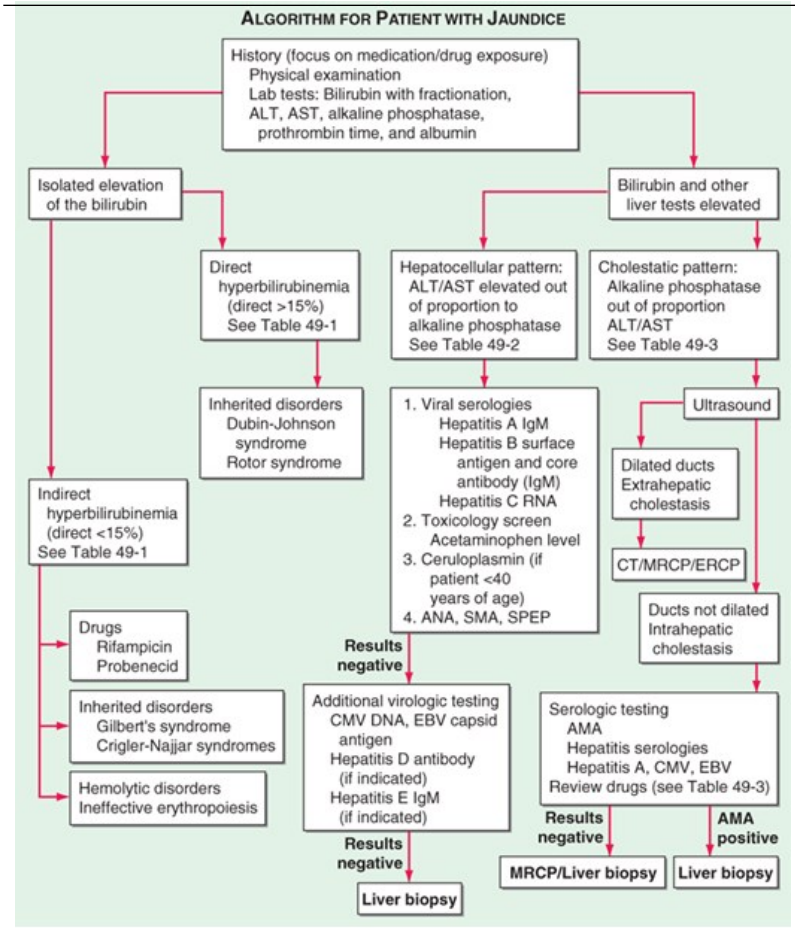
FIGURE 49-1

Evaluation of the patient with jaundice. ALT, alanine aminotransferase; AMA, antimitochondrial antibody; ANA, antinuclear antibody; AST, aspartate aminotransferase; CMV, cytomegalovirus; EBV, Epstein-Barr virus; ERCP, endoscopic retrograde cholangiopancreatography; LKM, liver-kidney microsomal antibody; MRCP, magnetic resonance cholangiopancreatography; SMA, smooth-muscle antibody; SPEP, serum protein electrophoresis.



Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

electrophoresis.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 49-1

Causes of Isolated Hyperbilirubinemia

I. Indirect hyperbilirubinemia
   A. Hemolytic disorders
   B. Ineffective erythropoiesis
   C. Increased bilirubin production
      1. Massive blood transfusion
      2. Resorption of hematoma
   D. Drugs
      1. Rifampin
      2. Probenecid
      3. Antibiotics—cephalosporins and penicilllins
   E. Inherited conditions
      1. Crigler-Najjar types I and II
      2. Gilbert's syndrome
II. Direct hyperbilirubinemia (inherited conditions)
   A. Dubin-Johnson syndrome
   B. Rotor syndrome

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Viral hepatitis

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 49-2

Hepatocellular Conditions That May Produce Jaundice

Viral hepatitis
Hepatitis A, B, C, D, and E
Epstein-Barr virus
Cytomegalovirus
Herpes simplex virus
Alcoholic hepatitis
Chronic liver disease and cirrhosis
Drug toxicity
Predictable, dose-dependent (e.g., acetaminophen)
Unpredictable, idiosyncratic (e.g., isoniazid)
Environmental toxins
Vinyl chloride
Jamaica bush tea—pyrrolizidine alkaloids
Kava kava
Wild mushrooms—*Amanita phalloides, A. verna*
Wilson's disease
Autoimmune hepatitis

TABLE 49-3

Cholestatic Conditions That May Produce Jaundice

I. Intrahepatic
   A. Viral hepatitis
      1. Fibrosing cholestatic hepatitis—hepatitis B and C
      2. Hepatitis A, Epstein-Barr virus infection, cytomegalovirus infection
   B. Alcoholic hepatitis
   C. Drug toxicity
      1. Pure cholestasis—anabolic and contraceptive steroids
      2. Cholestatic hepatitis—chlorpromazine, erythromycin estolate
      3. Chronic cholestasis—chlorpromazine and prochlorperazine

1. Chronic rejection of liver transplants
2. Sarcoidosis

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

B. Alcoholic hepatitis

3. Chronic cholestasis—chlorpromazine and prochlorperazine

D. Primary biliary cholangitis

E. Primary sclerosing cholangitis

F. Vanishing bile duct syndrome

1. Chronic rejection of liver transplants
2. Sarcoidosis
3. Drugs

G. Congestive hepatopathy and ischemic hepatitis

H. Inherited conditions

1. Progressive familial intrahepatic cholestasis
2. Benign recurrent intrahepatic cholestasis

I. Cholestasis of pregnancy

J. Total parenteral nutrition

K. Nonhepatobiliary sepsis

L. Benign postoperative cholestasis

M. Paraneoplastic syndrome

N. Veno-occlusive disease

O. Graft-versus-host disease

P. Infiltrative disease

1. Tuberculosis
2. Lymphoma
3. Amyloidosis

Q. Infections

1. Malaria
2. Leptospirosis

II. IExtrahepatic

A. Malignant

1. Cholangiocarcinoma
2. Pancreatic cancer
3. Gallbladder cancer
4. Ampullary cancer
5. Malignant involvement of the porta hepatis lymph nodes

B. Benign

1. Choledocholithiasis
2. Postoperative biliary strictures
3. Primary sclerosing cholangitis
4. Chronic pancreatitis
5. AIDS cholangiopathy
6. Mirizzi's syndrome
7. Parasitic disease (ascariasis)

GLOBAL CONSIDERATIONS

While extrahepatic biliary obstruction and drugs are common causes of new-onset jaundice in developed countries, infections remain the leading cause in developing countries. Liver involvement and jaundice are observed with numerous infections, particularly malaria, babesiosis, severe leptospirosis, infections due to *Mycobacterium tuberculosis* and the *Mycobacterium avium* complex, typhoid fever, infection with hepatitis viruses A–E, EBV, CMV, viral hemorrhagic fevers including Ebola virus, late phases of yellow fever, dengue fever, schistosomiasis, fascioliasis, clonorchiasis, opisthorchiasis, ascariasis, echinococcosis, hepatosplenic candidiasis, disseminated histoplasmosis, cryptococcosis, coccidioimycosis, ehrlichiosis, chronic Q fever, yersiniosis, brucellosis, syphilis, and leprosy. Bacterial infections that do not necessarily involve the liver and bile ducts may also lead to jaundice when they are severe enough to cause sepsis and multiorgan failure. Jaundice in patients with HIV infection is usually multifactorial and may be due to opportunistic infections, drugs, or neoplasms. Jaundice in a patient with HIV infection who is receiving highly active antiretroviral therapy (HAART) may be due to drug-induced hepatitis, which warrants urgent liver transplant evaluation.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACKNOWLEDGMENT

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

opisthorchiasis, ascariasis, echinococcosis, hepatosplenic candidiasis, disseminated histoplasmosis, cryptococcosis, coccidioimycosis, ehrlichiosis, chronic Q fever, yersiniosis, brucellosis, syphilis, and leprosy. Bacterial infections that do not necessarily involve the liver and bile ducts may also lead to jaundice, as in cholestasis of sepsis. The presence of fever or abdominal pain suggests concurrent infection, sepsis, or complications from gallstones. The development of encephalopathy and coagulopathy in a jaundiced patient with no preexisting liver disease signifies acute liver failure, which warrants urgent liver transplant evaluation.

ACKNOWLEDGMENT

This chapter is a revised version of chapters that have appeared in prior editions of Harrison's in which Marshall M. Kaplan was a co-author with Daniel Pratt.

FURTHER READING

Erlinger S et al: Inherited disorders of bilirubin transport and conjugation: New insights into molecular mechanisms and consequences. Gastroenterology 146:1625, 2014. [PubMed: 24704527]

Wolkoff AW et al: Bilirubin metabolism and jaundice, in *Schiff's Diseases of the Liver*, 11th ed, Schiff ER et al (eds). Oxford, UK, John Wiley & Sons, Ltd, 2012, pp 120–150.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 49: Jaundice, Savio John; Daniel S. Pratt
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 50: Abdominal Swelling and Ascites

Lawrence S. Friedman

ABDOMINAL SWELLING

Abdominal swelling is a manifestation of numerous diseases. Patients may complain of bloating or abdominal fullness and may note increasing abdominal girth on the basis of increased clothing or belt size. Abdominal discomfort is often reported, but pain is less frequent. When abdominal pain does accompany swelling, it is frequently the result of an intraabdominal infection, peritonitis, or pancreatitis. Patients with abdominal distention from *ascites* (fluid in the abdomen) may report the new onset of an inguinal or umbilical hernia. Dyspnea may result from pressure against the diaphragm and the inability to expand the lungs fully.

CAUSES

The causes of abdominal swelling can be remembered conveniently as the *six Fs*: flatus, fat, fluid, fetus, feces, or a "fatal growth" (often a neoplasm).

Flatus

Abdominal swelling may be the result of increased intestinal gas. The normal small intestine contains ~200 mL of gas made up of nitrogen, oxygen, carbon dioxide, hydrogen, and methane. Nitrogen and oxygen are consumed (swallowed), whereas carbon dioxide, hydrogen, and methane are produced intraluminally by bacterial fermentation. Increased intestinal gas can occur in a number of conditions. *Aerophagia*, the swallowing of air, can result in increased amounts of oxygen and nitrogen in the small intestine and lead to abdominal swelling. Aerophagia typically results from gulping food; chewing gum; smoking; or as a response to anxiety, which can lead to repetitive belching. In some cases, increased intestinal gas is the consequence of bacterial metabolism of excess fermentable substances such as lactose and other oligosaccharides, which can lead to production of hydrogen, carbon dioxide, or methane. In many cases, the precise cause of abdominal distention cannot be determined. In some persons, particularly those with irritable bowel syndrome and bloating, the subjective sense of abdominal pressure is attributable to impaired intestinal transit of gas rather than increased gas volume. Abdominal distention—an objective increase in girth—is the result of a lack of coordination between diaphragmatic contraction and anterior abdominal wall relaxation, a response in some cases to intraluminal bowel stimuli; dietary alterations, manipulation of the intestinal microbiota, or biofeedback may be effective therapy. Occasionally, increased lumbar lordosis accounts for apparent abdominal distention.

Fat

Weight gain with an increase in abdominal fat can result in an increase in abdominal girth and can be perceived as abdominal swelling. Abdominal fat may be caused by an imbalance between caloric intake and energy expenditure associated with a poor diet and sedentary lifestyle; it also can be a manifestation of certain diseases, such as Cushing's syndrome. Excess abdominal fat has been associated with an increased risk of insulin resistance and cardiovascular disease.

Fluid

The accumulation of fluid within the abdominal cavity (ascites) often results in abdominal distention and is discussed in detail below. Grade 1 ascites is detectable only by ultrasonography; grade 2 ascites is detectable by physical examination; and grade 3 ascites results in marked abdominal distention.

Fetus

Pregnancy results in increased abdominal girth. Typically, an increase in abdominal size is first noted at 12–14 weeks of gestation, when the uterus moves from the pelvis into the abdomen. Abdominal distention may be seen before this point as a result of fluid retention and relaxation of the abdominal muscles.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 50: Abdominal Swelling and Ascites, Lawrence S. Friedman

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

In the setting of severe constipation or intestinal obstruction, increased stool in the colon leads to increased abdominal girth. These conditions are often accompanied by abdominal discomfort or pain, nausea, and vomiting and can be diagnosed by imaging studies.

moves from the pelvis into the abdomen. Abdominal distention may be seen before this point as a result of fluid retention and relaxation of the abdominal muscles.

Feces

In the setting of severe constipation or intestinal obstruction, increased stool in the colon leads to increased abdominal girth. These conditions are often accompanied by abdominal discomfort or pain, nausea, and vomiting and can be diagnosed by imaging studies.

Fatal Growth

An abdominal mass can result in abdominal swelling. Neoplasms, abscesses, or cysts can grow to sizes that lead to increased abdominal girth. Enlargement of the intraabdominal organs, specifically the liver (hepatomegaly) or spleen (splenomegaly), or an abdominal aortic aneurysm can result in abdominal distention. Bladder distention also may result in abdominal swelling.

APPROACH TO THE PATIENT

Abdominal Swelling

HISTORY

Determining the etiology of abdominal swelling begins with history-taking and a physical examination. Patients should be questioned regarding symptoms suggestive of malignancy, including weight loss, night sweats, and anorexia. Inability to pass stool or flatus together with nausea or vomiting suggests bowel obstruction, severe constipation, or an ileus (lack of peristalsis). Increased eructation and flatus may point toward aerophagia or increased intestinal production of gas. Patients should be questioned about risk factors for or symptoms of chronic liver disease, including excessive alcohol use and jaundice, which suggest ascites. Patients should also be asked about symptoms of other medical conditions, including heart failure and tuberculosis, which may cause ascites.

PHYSICAL EXAMINATION

Physical examination should include an assessment for signs of systemic disease. The presence of lymphadenopathy, especially supraclavicular lymphadenopathy (*Virchow's node*), suggests metastatic abdominal malignancy. Care should be taken during the cardiac examination to evaluate for elevation of jugular venous pressure (JVP); *Kussmaul's sign* (elevation of the JVP during inspiration); a pericardial knock, which may be seen in heart failure or constrictive pericarditis; or a murmur of tricuspid regurgitation. Spider angiomas, palmar erythema, dilated superficial veins around the umbilicus (*caput medusae*), and gynecomastia suggest liver disease.

The abdominal examination should begin with inspection for the presence of uneven distention or an obvious mass. Auscultation should follow. The absence of bowel sounds or the presence of high-pitched localized bowel sounds points toward an ileus or intestinal obstruction. An umbilical venous hum may suggest the presence of portal hypertension, and a harsh bruit over the liver is heard rarely in patients with hepatocellular carcinoma or alcohol-associated hepatitis. Abdominal swelling caused by intestinal gas can be differentiated from swelling caused by fluid or a solid mass by percussion; an abdomen filled with gas is tympanic, whereas an abdomen containing a mass or fluid is dull to percussion. The absence of abdominal dullness, however, does not exclude ascites, because a minimum of 1500 mL of ascitic fluid is required for detection on physical examination. Finally, the abdomen should be palpated to assess for tenderness, a mass, enlargement of the spleen or liver, or presence of a nodular liver suggesting cirrhosis or tumor. Light palpation of the liver may detect pulsations suggesting retrograde vascular flow from the heart in patients with right-sided heart failure, particularly tricuspid regurgitation.

IMAGING AND LABORATORY EVALUATION

Abdominal x-rays can be used to detect dilated loops of bowel suggesting intestinal obstruction or ileus. Abdominal ultrasonography can detect as little as 100 mL of ascitic fluid, hepatosplenomegaly, a nodular liver, or a mass. Ultrasonography is often inadequate to detect retroperitoneal lymphadenopathy or a pancreatic lesion because of overlying bowel gas. If malignancy or pancreatic disease is suspected, CT can be performed. CT may also detect changes associated with advanced cirrhosis and portal hypertension **(Fig. 50-1)**.

FIGURE 50-1

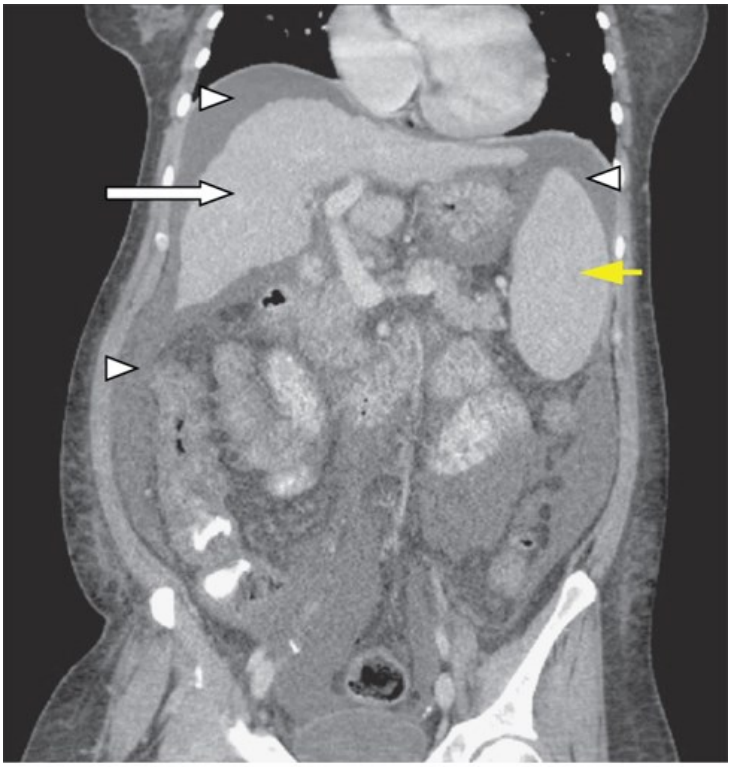
Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 50: Abdominal Swelling and Ascites, Lawrence S. Friedman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

FIGURE 50-1

CT of a patient with a cirrhotic, nodular liver (*white arrow*), splenomegaly (*yellow arrow*), and ascites (*arrowheads*). *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Laboratory evaluation should include liver biochemical testing, serum albumin level measurement, and prothrombin time determination (international normalized ratio) to assess hepatic function as well as a complete blood count to evaluate for the presence of cytopenias that may result from portal hypertension or of leukocytosis, anemia, and thrombocytosis that may result from systemic infection. Serum amylase and lipase levels should be checked to evaluate the patient for acute pancreatitis. Urinary protein quantitation is indicated when nephrotic syndrome, which may cause ascites, is suspected. Hydrogen and methane absorbed from the intestine are not metabolized by the host and are excreted in expired air, and detection of increased amounts of these gases in expired breath is the basis for tests used to diagnose carbohydrate (e.g., lactose) malabsorption and small intestinal bacterial overgrowth.

In selected cases, the hepatic venous pressure gradient (pressure across the liver between the portal and hepatic veins) can be measured via cannulation of the hepatic vein to confirm that ascites is caused by cirrhosis **(Chap. 344)**. In some cases, a liver biopsy may be necessary to confirm cirrhosis.

ASCITES

PATHOGENESIS IN THE PRESENCE OF CIRRHOSIS

Ascites in patients with cirrhosis is the result of portal hypertension and renal salt and water retention. Similar mechanisms contribute to ascites formation in heart failure. Portal hypertension signifies elevation of the pressure within the portal vein. According to Ohm's law, pressure is the product of resistance and flow. Increased hepatic resistance occurs by several mechanisms. First, the development of hepatic fibrosis, which defines cirrhosis, disrupts the normal architecture of the hepatic sinusoids and impedes normal blood flow through the liver. Second, activation of hepatic stellate cells, which mediate fibrogenesis, leads to smooth-muscle contraction and fibrosis. Finally, cirrhosis is associated with a decrease in endothelial nitric oxide synthetase (eNOS) production, which results in decreased nitric oxide production and increased intrahepatic vasoconstriction.

The development of cirrhosis is also associated with increased systemic circulating levels of nitric oxide (in contrast to the decrease seen

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 50: Abdominal Swelling and Ascites, Lawrence S. Friedman

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

perceived by the kidneys as hypovolemia. Compensatory vasoconstriction via release of antidiuretic hormone ensues; the consequences are free water retention and activation of the sympathetic nervous system and the renin-angiotensin-aldosterone system, which lead in turn to renal sodium and

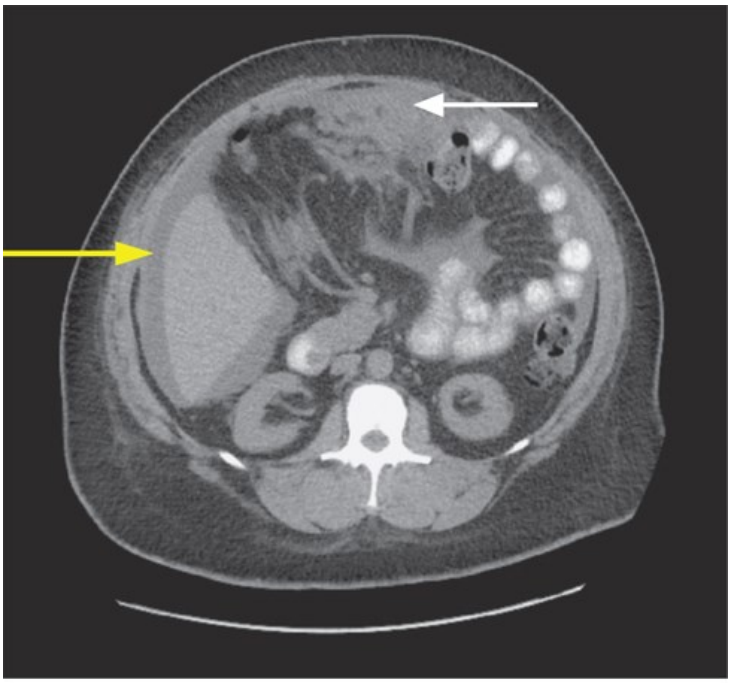
The development of cirrhosis is also associated with increased systemic circulating levels of nitric oxide (in contrast to the decrease seen intrahepatically), as well as increased levels of vascular endothelial growth factor and tumor necrosis factor, that result in splanchnic arterial vasodilation. Vasodilation of the splanchnic circulation results in pooling of blood and a decrease in the effective circulating volume, which is perceived by the kidneys as hypovolemia. Compensatory vasoconstriction via release of antidiuretic hormone ensues; the consequences are free water retention and activation of the sympathetic nervous system and the renin-angiotensin-aldosterone system, which lead in turn to renal sodium and water retention.

PATHOGENESIS IN THE ABSENCE OF CIRRHOSIS

Ascites in the absence of cirrhosis generally results from peritoneal carcinomatosis, peritoneal infection, or pancreatic disease. Peritoneal carcinomatosis can result from primary peritoneal malignancies such as mesothelioma or sarcoma, abdominal malignancies such as gastric or colonic adenocarcinoma, or metastatic disease from breast or lung carcinoma or melanoma **(Fig. 50-2)**. The tumor cells lining the peritoneum produce a protein-rich fluid that contributes to the development of ascites. Fluid from the extracellular space is drawn into the peritoneum, further contributing to the development of ascites. Tuberculous peritonitis causes ascites via a similar mechanism; tubercles deposited on the peritoneum exude a proteinaceous fluid. Pancreatic ascites results from leakage of pancreatic enzymes into the peritoneum.

FIGURE 50-2

CT of a patient with peritoneal carcinomatosis (*white arrow*) and ascites (*yellow arrow*). *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

CAUSES

Cirrhosis accounts for 84% of cases of ascites. Cardiac ascites, peritoneal carcinomatosis, and "mixed" ascites resulting from cirrhosis and a second disease account for 10–15% of cases. Less common causes of ascites include massive hepatic metastasis, infection (tuberculosis, *Chlamydia* infection), pancreatitis, and renal disease (nephrotic syndrome). Rare causes of ascites include hypothyroidism and familial Mediterranean fever.

EVALUATION

Once the presence of ascites has been confirmed, the etiology of the ascites is best determined by *paracentesis*, a bedside procedure in which a needle

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 50: Abdominal Swelling and Ascites, Lawrence S. Friedman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

infection, are infrequent.

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Once the presence of ascites has been confirmed, the etiology of the ascites is best determined by *paracentesis*, a bedside procedure in which a needle or small catheter is passed transcutaneously to extract ascitic fluid from the peritoneum. The lower quadrants are the most frequent sites for paracentesis. The left lower quadrant is preferred because of the greater depth of ascites and the thinner abdominal wall. Paracentesis is a safe procedure even in patients with coagulopathy; complications, including abdominal wall hematomas, hypotension, hepatorenal syndrome, and infection, are infrequent.

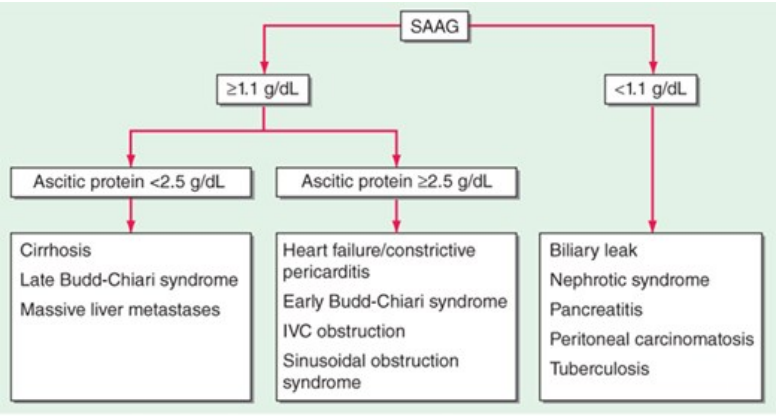
Once ascitic fluid has been extracted, its gross appearance should be examined. Turbid fluid can result from the presence of infection or tumor cells. White, milky fluid indicates a triglyceride level >200 mg/dL (and often >1000 mg/dL), which is the hallmark of *chylous ascites*. Chylous ascites results from lymphatic disruption that may occur with trauma, cirrhosis, tumor, tuberculosis, or certain congenital abnormalities. Dark brown fluid can reflect a high bilirubin concentration and indicates biliary tract perforation. Black fluid may indicate the presence of pancreatic necrosis or metastatic melanoma.

The ascitic fluid should be sent for measurement of albumin and total protein levels, cell and differential counts, and, if infection is suspected, Gram's stain and culture, with inoculation into blood culture bottles at the patient's bedside to maximize the yield. A serum albumin level should be measured simultaneously to permit calculation of the *serum-ascites albumin gradient* (SAAG).

The SAAG is useful for distinguishing ascites caused by portal hypertension from nonportal hypertensive ascites **(Fig. 50-3)**. The SAAG reflects the pressure within the hepatic sinusoids and correlates with the hepatic venous pressure gradient. The SAAG is calculated by subtracting the ascitic albumin concentration from the serum albumin level and does not change with diuresis. A SAAG ≥1.1 g/dL reflects the presence of portal hypertension and indicates that the ascites is due to increased pressure in the hepatic sinusoids. According to Starling's law, a high SAAG reflects the oncotic pressure that counterbalances the portal pressure. Possible causes include cirrhosis, cardiac ascites, hepatic vein thrombosis (Budd-Chiari syndrome), sinusoidal obstruction syndrome (veno-occlusive disease), or massive liver metastases. A SAAG <1.1 g/dL indicates that the ascites is not related to portal hypertension, as in tuberculous peritonitis, peritoneal carcinomatosis, or pancreatic ascites.

FIGURE 50-3

Algorithm for the diagnosis of ascites according to the serum-ascites albumin gradient (SAAG). IVC, inferior vena cava.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

For high-SAAG (≥1.1) ascites, the ascitic protein level can provide further clues to the etiology (Fig. 50-3). An ascitic protein level of ≥2.5 g/dL indicates that the hepatic sinusoids are normal and are allowing passage of protein into the ascites, as occurs in cardiac ascites, early Budd-Chiari syndrome, or sinusoidal obstruction syndrome. An ascitic protein level <2.5 g/dL indicates that the hepatic sinusoids have been damaged and scarred and no longer allow passage of protein, as occurs with cirrhosis, late Budd-Chiari syndrome, or massive liver metastases. Pro-brain-type natriuretic peptide (BNP) is a natriuretic hormone released by the heart as a result of increased volume and ventricular wall stretch. High levels of BNP in serum occur in heart failure and may be useful in identifying heart failure as the cause of high-SAAG ascites.

Further tests are indicated only in specific clinical circumstances. When secondary peritonitis resulting from a perforated hollow viscus is suspected, ascitic glucose and lactate dehydrogenase (LDH) levels can be measured. In contrast to "spontaneous" bacterial peritonitis, which may complicate cirrhotic ascites (see "Complications," below), secondary peritonitis is suggested by an ascitic glucose level <50 mg/dL, an ascitic LDH level higher than the serum LDH level, and the detection of multiple pathogens on ascitic fluid culture. When pancreatic ascites is suspected, the ascitic amylase level

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 50: Abdominal Swelling and Ascites, Lawrence S. Friedman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

without cirrhosis, an elevated ascitic adenosine deaminase level has a sensitivity of >90% for tuberculous ascites when a cut-off value of 30–45 U/L is used. When the cause of ascites remains uncertain, laparotomy or laparoscopy with peritoneal biopsies for histology and culture remains the gold

the serum LDH level, and the detection of multiple pathogens on ascitic fluid culture. When pancreatic ascites is suspected, the ascitic amylase level should be measured and is typically >1000 mg/dL. Cytology can be useful in the diagnosis of peritoneal carcinomatosis. At least 50 mL of fluid should be obtained and sent for immediate processing. Tuberculous peritonitis is typically associated with ascitic fluid lymphocytosis but can be difficult to diagnose by paracentesis. A smear for acid-fast bacilli has a diagnostic sensitivity of only 0–3%; a culture increases the sensitivity to 35–50%. In patients without cirrhosis, an elevated ascitic adenosine deaminase level has a sensitivity of >90% for tuberculous ascites when a cut-off value of 30–45 U/L is used. When the cause of ascites remains uncertain, laparotomy or laparoscopy with peritoneal biopsies for histology and culture remains the gold standard.

TREATMENT

Ascites

The initial treatment for cirrhotic ascites is restriction of sodium intake to 2 g/d. When sodium restriction alone is inadequate to control ascites, oral diuretics—typically the combination of spironolactone and furosemide—are used to increase urinary sodium excretion. Spironolactone is an aldosterone antagonist that inhibits sodium resorption in the distal convoluted tubule of the kidney. Use of spironolactone may be limited by hyponatremia, hyperkalemia, and painful gynecomastia. If the gynecomastia is distressing, amiloride (5–40 mg/d) may be substituted for spironolactone. Furosemide is a loop diuretic that is generally combined with spironolactone in a ratio of 40:100; maximal daily doses of spironolactone and furosemide are 400 mg and 160 mg, respectively. Fluid intake may be restricted in patients with hyponatremia.

Refractory cirrhotic ascites is defined by the persistence of ascites despite sodium restriction and maximal (or maximally tolerated) diuretic use. Pharmacologic therapy for refractory ascites includes the addition of midodrine, an α_1-adrenergic agonist, or clonidine, an α_2-adrenergic agonist, to diuretic therapy. These agents act as vasoconstrictors, counteracting splanchnic vasodilation. Midodrine alone or in combination with clonidine improves systemic hemodynamics and control of ascites over that obtained with diuretics alone. Although β-adrenergic blocking agents (beta blockers) are often prescribed to prevent variceal hemorrhage in patients with cirrhosis, the use of beta blockers in patients with refractory ascites may be associated with decreased survival rates.

When medical therapy alone is insufficient, refractory cirrhotic ascites can be managed by repeated large-volume paracentesis (LVP) or a transjugular intrahepatic peritoneal shunt (TIPS)—a radiologically placed portosystemic shunt that decompresses the hepatic sinusoids. Intravenous (IV) infusion of albumin accompanying LVP decreases the risk of "postparacentesis circulatory dysfunction" and death. Patients undergoing LVP should receive IV albumin infusions of 6–8 g/L of ascitic fluid removed. TIPS placement is superior to LVP in reducing the reaccumulation of ascites but is associated with an increased frequency of hepatic encephalopathy, with no difference in mortality rates. The Alfapump system, which consists of an automated pump and tunneled peritoneal catheter that transports ascites from the peritoneal cavity to the urinary bladder, has shown promise in the management of refractory ascites but is associated with a higher frequency of technical difficulties and renal dysfunction.

Malignant ascites does not respond to sodium restriction or diuretics. Patients must undergo serial LVPs, transcutaneous drainage catheter placement, or, rarely, creation of a peritoneovenous shunt (a shunt from the abdominal cavity to the vena cava) or placement of the Alfapump system, if available.

Ascites caused by tuberculous peritonitis is treated with standard antituberculosis therapy. Noncirrhotic ascites of other causes is treated by correction of the precipitating condition.

COMPLICATIONS

Spontaneous bacterial peritonitis (SBP; **Chap. 132**) is a common and potentially lethal complication of cirrhotic ascites. Occasionally, SBP also complicates ascites caused by nephrotic syndrome, heart failure, acute hepatitis, and acute liver failure but is rare in malignant ascites. Patients with SBP generally note an increase in abdominal girth; however, abdominal tenderness is found in only 40% of patients, and rebound tenderness is uncommon. Patients may present with fever, nausea, vomiting, or the new onset or an exacerbation of preexisting hepatic encephalopathy.

In hospitalized patients with ascites, paracentesis within 12 hours of admission reduces mortality because of early detection of SBP. SBP is defined by a polymorphonuclear neutrophil (PMN) count of ≥250/μL in the ascitic fluid. Cultures of ascitic fluid should be performed in blood culture bottles and typically reveal one bacterial pathogen. The presence of multiple pathogens in the setting of an elevated ascitic PMN count suggests *secondary*

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 50: Abdominal Swelling and Ascites, Lawrence S. Friedman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Treatment of SBP with an antibiotic such as IV cefotaxime is generally effective against gram-negative and gram-positive aerobes. A 5-day course of

polymorphonuclear neutrophil (PMN) count of ≥250/μL in the ascitic fluid. Cultures of ascitic fluid should be performed in blood culture bottles and typically reveal one bacterial pathogen. The presence of multiple pathogens in the setting of an elevated ascitic PMN count suggests *secondary peritonitis* from a ruptured viscus or abscess **(Chap. 132)**. The presence of multiple pathogens without an elevated PMN count suggests bowel perforation from the paracentesis needle. SBP is generally the result of enteric bacteria that have translocated across an edematous bowel wall. The most common pathogens are gram-negative rods, including *Escherichia coli* and *Klebsiella*, as well as streptococci and enterococci.

Treatment of SBP with an antibiotic such as IV cefotaxime is generally effective against gram-negative and gram-positive aerobes. A 5-day course of treatment is sufficient if the patient improves clinically. Nosocomial or health care–acquired SBP is frequently caused by multidrug-resistant bacteria, and initial antibiotic therapy should be guided by the local bacterial epidemiology.

Cirrhotic patients with a history of SBP, an ascitic fluid total protein concentration <1 g/dL, or active gastrointestinal bleeding should receive prophylactic antibiotics to prevent SBP; oral daily ciprofloxacin or, where available, norfloxacin is commonly used. IV ceftriaxone may be used in hospitalized patients. Diuresis increases the activity of ascitic fluid protein opsonins and may decrease the risk of SBP.

Hepatic hydrothorax occurs when ascites, often caused by cirrhosis, migrates via fenestrae in the diaphragm into the pleural space. This condition can result in shortness of breath, hypoxia, and infection. Treatment is similar to that for cirrhotic ascites and includes sodium restriction, diuretics, and, if needed, thoracentesis or TIPS placement. Chest tube placement should be avoided.

ACKNOWLEDGMENT

The author thanks Dr. Kathleen E. Corey for contributions to this chapter in prior editions of the textbook.

FURTHER READING

Adebayo D et al: Refractory ascites in liver cirrhosis. Am J Gastroenterol 114:40, 2019. [PubMed: 29973706]

Barba E et al: Correction of abdominal distention by biofeedback-guided control of abdominothoracic muscular activity in a randomized, placebo-controlled trial. Clin Gastroenterol Hepatol 15:1922, 2017. [PubMed: 28705783]

Bernardi M et al: Albumin infusion in patients undergoing large-volume paracentesis: A meta-analysis of randomized trials. Hepatology 55:1172, 2012. [PubMed: 22095893]

European Association for the Study of the Liver: EASL Clinical Practice Guidelines for the management of patients with decompensated cirrhosis. J Hepatol 69:406, 2018. [PubMed: 29653741]

Farias AQ et al: Serum B-type natriuretic peptide in the initial workup of patients with new onset ascites: A diagnostic accuracy study. Hepatology 59:1043, 2014. [PubMed: 23907731]

Fernandez J et al: Prevalence and risk factors of infections by multiresistant bacteria in cirrhosis: A prospective study. Hepatology 55:1551, 2012. [PubMed: 22183941]

Ge PS, Runyon BA: Role of plasma BNP in patients with ascites: Advantages and pitfalls. Hepatology 59:751, 2014. [PubMed: 23959954]

John S, Friedman LS: Portal hypertensive ascites: Current status. Curr Hepatol Rep 19:226, 2020.

John S, Thuluvath PJ: Hyponatremia in cirrhosis: Pathophysiology and management. World J Gastroenterol 21:3197, 2015. [PubMed: 25805925]

Lizaola B et al: Review article: the diagnostic approach and current management of chylous ascites. Aliment Pharmacol Ther 46:816, 2017. [PubMed: 28892178]

Malagelada JR et al: Bloating and abdominal distension: Old misconceptions and current knowledge. Am J Gastroenterol 112:1221, 2017. [PubMed: 28508867]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 50: Abdominal Swelling and Ascites, Lawrence S. Friedman

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Runyon BA: Introduction to the revised American Association for the Study of Liver Diseases Practice Guideline management of adult patients with

ACCESS Medicine

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Orman ES et al: Paracentesis is associated with reduced mortality in patients hospitalized with cirrhosis and ascites. Clin Gastroenterol Hepatol 12:496, 2014. [PubMed: 23978348]

Runyon BA: Introduction to the revised American Association for the Study of Liver Diseases Practice Guideline management of adult patients with ascites due to cirrhosis 2012. Hepatology 57:165, 2013.

Runyon BA et al: The serum-ascites albumin gradient is superior to the exudate-transudate concept in the differential diagnosis of ascites. Ann Intern Med 117:215, 1992. [PubMed: 1616215]

Sort P et al: Effect of intravenous albumin on renal impairment and mortality in patients with cirrhosis and spontaneous bacterial peritonitis. N Engl J Med 341:403, 1999. [PubMed: 10432325]

Williams JW jr, Simel DL: The rational clinical examination. Does this patient have ascites? How to divine fluid in the abdomen. JAMA 267:2645, 1992. [PubMed: 1573754]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 50: Abdominal Swelling and Ascites, Lawrence S. Friedman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome

R. Christopher Doiron; J. Curtis Nickel

DEFINITION

A condition associated with bladder inflammation and pain, with what were thought to be discrete bladder ulcerations, was first described in 1887. The description of the classic bladder-wall ulcer—now referred to as a *Hunner lesion*—became known as *interstitial cystitis* (IC). The first generally accepted definition of IC was derived from a National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) consensus of experts in the field in 1998. The NIDDK criteria used to define IC included typical cystoscopic findings such as glomerulations (submucosal petechial hemorrhages of the urothelium) or Hunner lesions. However, over time, the syndrome experienced by patients, including bladder and/or pelvic pain with associated urinary storage symptoms of urinary frequency and urgency, negative urine cultures, and no specific identifiable causes, became known as *interstitial cystitis/bladder pain syndrome* (IC/BPS).

The nomenclature and definitions have evolved, but the contemporary definitions accepted by the American Urological Association, the Canadian Urological Association, the International Continence Society, the Society for Urodynamics and Female Urology, and the European Society for the Study of IC/BPS, although they all differ somewhat in language and specifics, generally reflect several fundamental concepts common in the disease: (1) it is chronic in nature; (2) it causes pain perceived to be attributable to the bladder; (3) this pain occurs in the presence of lower urinary tract symptoms (LUTS); and (4) pain outside the bladder—in the pelvis, perineum, genitals, abdomen, and beyond—is common.

The following definition incorporates the major descriptions by all international groups interested in the diagnosis and management of IC/BPS: an unpleasant sensation (pain, pressure, discomfort) perceived to be related to the urinary bladder, associated with LUTS of >6 weeks' duration, in the absence of infection or other identifiable causes.

A generalized urologic chronic pelvic pain syndrome (UCPPS) is referenced in the literature and is thought to encompass two distinct urologic chronic pain disorders: IC/BPS, which may be present in men and women, and chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS), which is present only in men. The latter refers to a urologic pain disorder with pain localized to the perineum and/or male genitals, with or without LUTS. IC/BPS can exist independent of CP/CPPS in men. In reality, the urologic chronic pain disorders often have overlapping symptom presentations and may share common etiologic and pathophysiologic origins, but the focus of the current chapter will be on IC/BPS.

ETIOLOGY AND PATHOGENESIS

Pinning a single etiology to a diagnosis of IC/BPS has been an endeavor fraught with uncertainty that ultimately has failed thus far. Instead, it is much more likely and widely accepted that IC/BPS represents a syndrome or constellation of interrelated disease processes that manifest in a spectrum of disease that reaches beyond the bladder. While the search for a single etiology soldiers on, we will review here a collection of proposed theories.

INFECTION AND THE URINARY MICROBIOTA

Bacterial infection of the urothelium has long been regarded as a major suspect in the etiology of IC/BPS but has never been definitively shown to cause the disease. It is not uncommon for patients presenting with IC/BPS to describe a long history of "urinary tract infections" (UTIs); these patients often have undergone multiple courses of treatment with one or multiple antibiotics prescribed by their physicians. Often, however, in patients with IC/BPS, the benefit of antibiotic treatment is short-lived, urine culture results are negative, and the return of symptoms is inevitable.

Although studies examining the role of microbiologic organisms in this patient population are numerous and the results conflicting, far more studies have yielded negative results rather than positive findings. Furthermore, our understanding of the urinary microbiota continues to expand, rendering older studies using outdated and insensitive cultivation techniques less relevant.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome, R. Christopher Doiron; J. Curtis Nickel

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

clinicians believe that although a single causative microbe is unlikely, dysbiosis or disturbance in the microbial ecology of the lower urinary tract may be responsible for flares or symptom patterns experienced by IC/BPS patients.

older studies using outdated and insensitive cultivation techniques less relevant.

Using state-of-the-art, culture-independent techniques for microorganism identification, investigators observe subtle differences between the urinary microbiota of IC/BPS patients and that of healthy controls and between IC/BPS patients experiencing symptom flares and IC/BPS patients not in flare. The clinical relevance of these findings is still not fully understood. As the study of the urinary microbiota continues to unfold, researchers and clinicians believe that although a single causative microbe is unlikely, dysbiosis or disturbance in the microbial ecology of the lower urinary tract may be responsible for flares or symptom patterns experienced by IC/BPS patients.

AUTOIMMUNITY

The consideration of IC/BPS as a disorder of the immune system stems from the observation of a significant prevalence of autoimmune disorders in IC/BPS patients; several historical studies have identified anti-urothelial antibodies within the bladder mucosa of IC/BPS patients. Furthermore, although IC/BPS is not a pathologic diagnosis, there are widely accepted, recognizable patterns of inflammatory infiltration in the bladder mucosa of this patient population, including lymphoplasmacytic infiltrates, stromal edema and fibrosis, urothelial denudation, and detrusor mastocytosis. Thus, although it is likely that immune disturbances cause the condition in a subset of patients (for example, in those with associated Sjögren's syndrome), researchers and clinicians have been unable to leverage this knowledge into a clear description, and its clinical relevance is not fully understood.

INFLAMMATION

It is well established that a subset of patients suffering from IC/BPS clearly have associated bladder inflammation of unknown etiology. The best described of these patients are those with Hunner lesions—discrete inflammatory lesions, previously believed to be ulcers, that have a well-characterized inflammatory profile on histologic and pathologic analysis. While Hunner lesions are easily identified under direct vision by cystoscopy, a spectrum of other, less obvious inflammatory patterns in the bladder is associated with infiltration of acute and chronic inflammatory cells and mast cells. This inflammation observed on histologic analysis can be so subtle that it cannot be recognized under direct visual examination of the bladder with cystoscopy.

Investigators in the Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP) Research Network have found that, among patients with UCPPS, women exhibit more robust inflammatory responses to stimulation of Toll-like receptor 2 (TLR2) and TLR4. Furthermore, an increased response to stimulation of TLR4 predicts more severe symptoms, widespread pain (vs pelvic/bladder pain only), and a higher number of chronic overlapping pain conditions (COPCs). Further studies aimed at a better understanding of these findings are under way.

UROTHELIAL DYSFUNCTION

Urothelial Permeability and the Glycosaminoglycan Layer

The stratified epithelium of the bladder—the urothelium—is composed of basal precursor cells, intermediate cells, and a layer of specialized, superficial epithelial cells called *umbrella cells*. Collectively, these layers are responsible for the various functions of the bladder lining. One important function of the urothelium is to provide a robust barrier layer. This function is fulfilled by the dense layering of glycosaminoglycans (GAGs) on the luminal surface of the urothelium along with a complex arrangement of numerous intercellular tight junctions among urothelial cells that protects the underlying bladder interstitium from the constituents of the urine resting in the bladder.

Defects in this barrier function—either disruptions in the GAG layer or disruptions in the epithelial layer itself or its cellular junctions—have been proposed as a possible mechanism for bladder pain in IC/BPS patients. This theory, while still popular, lacks definitive evidence supporting it as the etiology of this disease.

Antiproliferative Factor

The discovery that urothelial cells from IC/BPS patients appear to grow far more slowly than urothelial cells from a healthy control population led to the identification of antiproliferative factor (APF). Although APF initially showed promise as a sensitive and specific urine biomarker for IC/BPS, this idea has not been widely adopted, and the etiologic role of APF is not yet fully understood.

PELVIC ORGAN CROSSTALK

The observation of dysfunction and symptoms in multiple organ systems, including gastrointestinal, gynecologic, and genital organs, in patients with

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

NEUROBIOLOGIC CONTRIBUTIONS AND CENTRAL SENSITIZATION

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

The observation of dysfunction and symptoms in multiple organ systems, including gastrointestinal, gynecologic, and genital organs, in patients with IC/BPS is so common that it might be considered the norm. Mechanisms of neural sensitization in patients with chronic pain have been reported, and abnormalities in the autonomic nervous system have been observed among IC/BPS patients. Again, although these observations apply in a subset of patients, their broader application to the heterogeneous IC/BPS patient population as a clear cause of disease is not warranted.

NEUROBIOLOGIC CONTRIBUTIONS AND CENTRAL SENSITIZATION

One breakthrough by the MAPP Research Network is an investigation of the role of structural and functional alterations in the brains of patients with UCPPS. The network's innovative methods of correlating clinical and deep phenotyping data with functional MRI data identified such structural and functional differences. These differences were later shown to successfully predict the progression of symptoms in a cohort of 52 patients with UCPPS. Although the relevant study did not differentiate between IC/BPS and CP/CPPS patients, the findings are nevertheless informative, and further longitudinal studies are ongoing.

In addition to the novel MAPP-led findings using neuroimaging, quantitative sensory testing (QST) methods have been used to investigate the sensory processing mechanisms in UCPPS patients. The findings—generalized pain hypersensitivity and altered endogenous inhibitory pain control systems among UCPPS patients—further support a hypothesis of a central sensitization phenotype in urologic chronic pelvic pain. The clinical implications of observed neural alterations and multisensory hypersensitivity remain under investigation.

Although a single etiology for this clinically heterogeneous pain syndrome may never be identified, efforts to do so have revealed much about its pathogenesis in subsets of patients and have provided valuable insight into specific patient phenotypes. The challenge for researchers and clinicians moving forward will be to unify clinical phenotypes with these proposed underlying mechanisms of disease and to integrate this knowledge into clinically actionable interventions that may provide meaningful outcomes.

EPIDEMIOLOGY

The prevalence of IC/BPS has been difficult to determine because definitions and diagnostic criteria (in the absence of a definitive diagnostic test or biomarker) are constantly evolving. In addition, the various methods used in attempting to describe the syndrome's epidemiology (patient self-reports, symptom-based surveys, physician visits, population-based databases) have been problematic and have made comparisons of results challenging. Many studies have historically been performed only in female populations. Currently, it is estimated that 2.7–6.5% of North American women experience symptoms consistent with a diagnosis of IC/BPS. Fewer than 10% of women who experience these symptoms actually have a diagnosis of IC/BPS. The syndrome does occur in men, with a reported 10:1 female-to-male ratio, but it is thought that the condition is dramatically underreported in men.

Some predictors of the development of IC/BPS have been suggested through an analysis of retrospective observational studies of childhood disorders and adverse childhood experiences (ACEs), including childhood UTI, childhood bowel and bladder dysfunction, and childhood sexual trauma. Furthermore, it has been well established that IC/BPS patients exhibit a remarkable prevalence of COPCs such as fibromyalgia, irritable bowel syndrome (IBS), chronic back pain, and chronic fatigue syndrome (CFS). Recent MAPP-led studies have shown that more than one-third of IC/BPS patients have one COPC (IBS, fibromyalgia, or CFS), while up to 10% have multiple COPCs. Thus, these conditions might be considered risk factors for the development of IC/BPS.

CLINICAL MANIFESTATIONS

Patients with IC/BPS, both female and male, present with varying degrees of discomfort and/or pain perceived to be related to the bladder and associated with urinary storage symptoms, including daytime and nighttime urinary frequency and urinary urgency. For some patients, urinary symptoms (the most common complaint after bladder pain) are the most bothersome, while for most patients, bladder pain causes the most distress and most significantly affects quality of life. Unfortunately, the majority of patients with IC/BPS present with both types of symptoms, as patients void frequently to relieve pain (or because of fear of bladder pain). Typically, this combination of bladder pain and urinary frequency severely impacts patients' quality of life, social interactions, and physical activities.

Pain-mapping studies have been used to identify different pain phenotypes within the disease. Nickel and colleagues first described a bladder-only phenotype present in 20% of a cohort of female IC/BPS patients, whereas up to 80% of patients described pain in the pelvis and at least one site beyond. Common associated conditions include IBS (40%), pelvic floor dysfunctional pain syndrome (40–60%), vulvodynia (17%), fibromyalgia (36%), CFS (10%), and chronic back pain (47%). As described above, these multiorgan symptoms may be due to central nervous system sensitization and a pelvic pain–only phenotype in 25% of participants and pain in the pelvis and beyond in up to 75%.

Another important finding from the MAPP investigations is their identification of not just a pelvic-pain-only phenotype but also of a bladder-focused

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome, R. Christopher Doiron; J. Curtis Nickel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

beyond. Common associated conditions include IBS (40%), pelvic floor dysfunctional pain syndrome (40–60%), vulvodynia (17%), fibromyalgia (36%), CFS (10%), and chronic back pain (47%). As described above, these multiorgan symptoms may be due to central nervous system sensitization and associated spinal crosstalk, which may promote phenotypic progression as patients with one pain syndrome slowly progress to another. Subsequent MAPP-led studies among a more heterogeneous UCPPS cohort of men and women have supported this concept of specific pain phenotypes, reporting a pelvic-pain-only phenotype in 25% of participants and pain in the pelvis and beyond in up to 75%.

Another important finding from the MAPP investigations is their identification of not just a pelvic-pain-only phenotype but also of a bladder-focused phenotype. The latter phenotype was identified by patients' responses to two RAND Interstitial Cystitis Epidemiology (RICE) survey questions: whether they had "painful bladder filling" and/or "painful urinary urgency." Most female UCPPS patients (88%) responded "yes" to at least one of these questions. The bladder-focused phenotype was associated with more severe urologic symptoms and worse quality of life.

Patients present with unique pain trajectories. Some initially have mild discomfort that progresses over many years to pain with bladder filling and finally to chronic unremitting pelvic pain with only short periods of relief with urination. Other patients begin with UTI-like symptoms and acute bladder and urethral pain with urinary frequency and urgency; these manifestations persist as a chronic cystitis–like syndrome despite negative cultures and no benefit from antimicrobial therapy. Still other patients report a waxing and waning of pain over time, with flares exacerbated by diet, anxiety/stress, infection, or hormone cycle (typically with increased pain prior to menses). In a longitudinal study of UCPPS patients followed over a 12-month period during routine care for their disease, MAPP investigators described 60% of patients' symptoms as stable, 20% as improved, and 20% as worsened.

APPROACH TO THE PATIENT

Interstitial Cystitis/Bladder Pain Syndrome

Patients with IC/BPS present to their family physician or internist with pelvic pain that typically increases in severity with bladder filling, other associated pain, and various degrees of urinary symptomatology. The course that should be followed by primary care practitioners or general internists during the patient's workup and before referral to a specialist is outlined in **Table 51-1**. Most of these physicians will not move beyond a suspected diagnosis and conservative advice; that is acceptable. Patients with IC/BPS can often represent diagnostic challenges, and referral to an appropriate subspecialist is warranted if any diagnostic uncertainty remains.

A diagnosis of IC/BPS is often missed and delayed for many years because physicians tend to silo patients into various medical-specialty streams on the basis of the predominant or most bothersome symptom. For example, patients presenting with pelvic pain in which flares are associated with monthly menstrual cycles may be referred to gynecologists. Patients with abdominal/pelvic pain associated with diarrhea and/or constipation tend to be referred to gastroenterologists, while those with generalized muscle and joint pain, perhaps associated with fatigue, are referred to rheumatologists. Patients with urinary symptoms and bladder pain are treated for UTIs (even with negative urine cultures) or overactive bladder—a common bladder condition associated with urinary frequency and urgency, but not pain.

It would be simple if the approach to patients presenting with pelvic pain was only to determine the actual pelvic organ and/or disease causing the symptoms. However, spinal crosstalk, phenotype progression over time, central sensitization, and COPCs complicate the picture. Since only ~20–25% of patients eventually diagnosed with IC/BPS have bladder-only disease, one must not be bladder-centric in approach but rather must consider the entire patient. The provider must determine the patient's "clinical picture"—that is, the patient's unique presenting clinical phenotype.

Urologists have adapted a system of clinical symptom categorization for patients with UCPPS. UPOINT, which includes documenting the contribution of six distinct domains—*U*rinary, *P*sychosocial, *O*rgan-specific, *I*nfection, *N*eurologic, and *T*enderness (as in pelvic floor muscle tenderness)—has helped categorize patient symptoms and allows the practitioner to focus their management on the most bothersome domain, while helping to avoid neglecting domains that are often forgotten. While used by many urologists managing this condition, UPOINT is not as effective in IC/BPS as it is in male CP/CPPS, probably because all IC/BPS patients would be categorized, by definition, in the U and O domains.

A further simplified clinical approach to the assessment of patients with symptoms of IC/BPS is to classify patients with perceived bladder pain (a mandatory criterion for diagnosis) into one of two categories: (1) a "pelvic-pain-only" category, which would include the "bladder-pain-only," pelvic floor dysfunctional pain, and associated gynecologic pain groups; or (2) a "pelvic pain and beyond" category, which would include patients with associated COPCs (such as IBS and fibromyalgia). This approach has been supported by recent observations from the MAPP investigators.

The contribution of psychosocial parameters, such as depression, catastrophizing, anxiety, and stress, and their impact on pain and disability cannot be overlooked and are important to ascertain in all cases. This approach to clinical phenotyping will let the physician tailor a unique

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 51-1

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

APPROACH TO THE PATIENT

Interstitial Cystitis/Bladder Pain Syndrome

Patients with IC/BPS present to their family physician or internist with pelvic pain that typically increases in severity with bladder filling, other associated pain, and various degrees of urinary symptomatology. The course that should be followed by primary care practitioners or general internists during the patient's workup and before referral to a specialist is outlined in **Table 51-1**. Most of these physicians will not move beyond a suspected diagnosis and conservative advice; that is acceptable. Patients with IC/BPS can often represent diagnostic challenges, and referral to an appropriate subspecialist is warranted if any diagnostic uncertainty remains.

A diagnosis of IC/BPS is often missed and delayed for many years because physicians tend to silo patients into various medical-specialty streams on the basis of the predominant or most bothersome symptom. For example, patients presenting with pelvic pain in which flares are associated with monthly menstrual cycles may be referred to gynecologists. Patients with abdominal/pelvic pain associated with diarrhea and/or constipation tend to be referred to gastroenterologists, while those with generalized muscle and joint pain, perhaps associated with fatigue, are referred to rheumatologists. Patients with urinary symptoms and bladder pain are treated for UTIs (even with negative urine cultures) or overactive bladder—a common bladder condition associated with urinary frequency and urgency, but not pain.

It would be simple if the approach to patients presenting with pelvic pain was only to determine the actual pelvic organ and/or disease causing the symptoms. However, spinal crosstalk, phenotype progression over time, central sensitization, and COPCs complicate the picture. Since only ~20–25% of patients eventually diagnosed with IC/BPS have bladder-only disease, one must not be bladder-centric in approach but rather must consider the entire patient. The provider must determine the patient's "clinical picture"—that is, the patient's unique presenting clinical phenotype.

Urologists have adapted a system of clinical symptom categorization for patients with UCPPS. UPOINT, which includes documenting the contribution of six distinct domains—*U*rinary, *P*sychosocial, *O*rgan-specific, *I*nfection, *N*eurologic, and *T*enderness (as in pelvic floor muscle tenderness)—has helped categorize patient symptoms and allows the practitioner to focus their management on the most bothersome domain, while helping to avoid neglecting domains that are often forgotten. While used by many urologists managing this condition, UPOINT is not as effective in IC/BPS as it is in male CP/CPPS, probably because all IC/BPS patients would be categorized, by definition, in the U and O domains.

A further simplified clinical approach to the assessment of patients with symptoms of IC/BPS is to classify patients with perceived bladder pain (a mandatory criterion for diagnosis) into one of two categories: (1) a "pelvic-pain-only" category, which would include the "bladder-pain-only," pelvic floor dysfunctional pain, and associated gynecologic pain groups; or (2) a "pelvic pain and beyond" category, which would include patients with associated COPCs (such as IBS and fibromyalgia). This approach has been supported by recent observations from the MAPP investigators.

The contribution of psychosocial parameters, such as depression, catastrophizing, anxiety, and stress, and their impact on pain and disability cannot be overlooked and are important to ascertain in all cases. This approach to clinical phenotyping will let the physician tailor a unique treatment plan for each individual patient, using combinations of local bladder, pelvic floor, or more general systemic therapies.

TABLE 51-1

Workup of Patients by a Primary Care Practitioner or General Internist

STEPS IN WORKUP	SPECIFICS
History/physical examination	Conduct a pelvic exam (recommended). Categorize symptoms as bladder/pelvis focused and/or extending beyond the pelvis.
Urinalysis	Perform a urine culture. If the culture is positive, conduct sensitivity testing.
Consideration of patient-centered treatment options if satisfied with diagnosis	Begin with conservative measures. Introduce further symptom-specific treatments as needed
	• the diagnosis is unclear • microscopic or gross hematuria is present

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome, R. Christopher Doiron; J. Curtis Nickel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 51-1

Workup of Patients by a Primary Care Practitioner or General Internist

STEPS IN WORKUP	SPECIFICS
History/physical examination	Conduct a pelvic exam (recommended). Categorize symptoms as bladder/pelvis focused and/or extending beyond the pelvis.
Urinalysis	Perform a urine culture. If the culture is positive, conduct sensitivity testing.
Consideration of patient-centered treatment options if satisfied with diagnosis[a]	Begin with conservative measures. Introduce further symptom-specific treatments as needed.
Referral to an appropriate specialist under certain conditions	Referral should follow if: • the diagnosis is unclear • microscopic or gross hematuria is present • the condition is refractory to treatment • symptoms are severe • the presentation is complex

[a]See text.

DIAGNOSIS

IC/BPS is a clinically heterogeneous condition whose lack of a clear etiopathogenesis presents difficulties in diagnosis. In making a diagnosis of exclusion, clinicians must rule out other confusable diseases and identify to the best of their ability the phenotypic presentation of the presenting patient. Although attempts have been made to establish a set of diagnostic criteria in the past, the specified criteria have proven overly stringent and too exclusive to be clinically useful. Furthermore, although several guidelines exist to aide in decision making in diagnostic investigations, most investigations serve merely to rule out other pathology. In contrast, history and physical examination, along with some simple laboratory testing, are the most reliable tools with which to establish a diagnosis of IC/BPS. Details of relevant investigations, some of which may be beyond the scope of the general practitioner or internist, are presented here. Table 51-1 offers an approach for the general practitioner, and **Table 51-2** provides a more complete summary of diagnostic recommendations.

TABLE 51-2

Recommendations for Investigations in Patients with Suspected Interstitial Cystitis/Bladder Pain Syndrome

MANDATORY	RECOMMENDED	OPTIONAL	NOT RECOMMENDED
History	Frequency/volume chart	Ultrasound/pelvic imaging	Potassium sensitivity test
Physical examination	Urinalysis	Postvoid residual	Urodynamics
	Urine culture	Urine cytology	Bladder biopsy
	Symptom scores	Intravesical anesthetic bladder challenge	

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome, R. Christopher Doiron; J. Curtis Nickel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Source: Adapted from A Cox et al: CUA guideline: Diagnosis and treatment of interstitial cystitis/bladder pain syndrome. Can Urol Assoc J 10:E136, 2016.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 51-2

Recommendations for Investigations in Patients with Suspected Interstitial Cystitis/Bladder Pain Syndrome

MANDATORY	RECOMMENDED	OPTIONAL	NOT RECOMMENDED
History	Frequency/volume chart	Ultrasound/pelvic imaging	Potassium sensitivity test
Physical examination	Urinalysis	Postvoid residual	Urodynamics
	Urine culture	Urine cytology	Bladder biopsy
	Symptom scores	Intravesical anesthetic bladder challenge	
	Cystoscopy	Hydrodistension	

Source: Adapted from A Cox et al: CUA guideline: Diagnosis and treatment of interstitial cystitis/bladder pain syndrome. Can Urol Assoc J 10:E136, 2016.

HISTORY AND PHYSICAL (INCLUDING FREQUENCY/VOLUME CHARTS)

A thorough history and physical examination are of utmost importance in diagnosing IC/BPS. A history of the patient's pain symptoms is a logical place to start. The nature, intensity, and timing of the pain are all significant factors. Some patients will be less explicit than others in describing their pain and may instead describe a sense of pressure, burning, or vague fullness in the pelvis or bladder area.

All aspects of the patient's pain should be explored, as many patients' pain will not be limited to the pelvis or bladder but will be associated with the genitals, anus or rectum, perineum, abdomen, and beyond. Furthermore, although pain is commonly experienced with bladder filling, patients may also have suprapubic tenderness or pressure with voiding or burning or pain in the bladder, urethra, or perineum, with radiation into the vagina for women or the prostate, penis, and testicles for men. In male patients, distinguishing IC/BPS from CP/CPPS can be challenging. Physicians must assess for more widespread pain locations outside the pelvis; screening for COPCs, particularly IBS, fibromyalgia, CFS, back pain, and headache, is important in adequately addressing the clinical impact of IC/BPS.

Eliciting and understanding associated LUTS—specifically urinary frequency, urgency, and nocturia—should be another focus of the history. While several confusable diseases can present with LUTS, the manifestation of IC/BPS as voiding dysfunction can help guide treatment decisions and is often a significant focus of bother for the patient. Having patients complete frequency/volume charts, noting the time and volume of each urination over a 24-hour period, can help provide objective evidence of a LUTS history and facilitate follow-up during and after treatment.

Physical examination should focus on the abdomen, pelvis, genitals, and pelvic floor. The degree of pelvic floor relaxation (i.e., degree of muscle tension and/or spasm) during examination is important to note. Trigger points in the pelvic floor musculature and any areas of localized spasticity should be identified. In women, an examination of the vulva, vaginal mucosa, and urethral meatus is essential to identify the presence of vulvodynia (vulvar mucosal pain with no identifiable cause) or any signs of genitourinary syndrome of menopause. In men, an examination of the external genitalia and a digital rectal (prostate) examination as well as a similar pelvic floor examination should be included to rule out related pathology.

SYMPTOM SCORES

The quality of a history can be elevated by an accompanying validated, objective measurement of the patient's symptoms. Although several relevant tools exist, the Interstitial Cystitis Symptom Index (ICSI) and the Interstitial Cystitis Problem Index (ICPI) are the most widely used, are commonly employed in research trials as outcome measures, and are straightforward enough for the practitioner to perform in an outpatient setting. These short questionnaires document pain severity, urinary frequency, urgency, and nocturia as well as the bother experienced from each of these symptoms.

More recently, MAPP investigators have suggested that pain and urinary symptoms should be assessed independently using two separate questionnaires: the Genitourinary Pain Index (GUPI) to assess pain and the ICSI to separately assess urinary symptoms. This suggestion is based on

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome, R. Christopher Doiron; J. Curtis Nickel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

URINE STUDIES

questionnaires: the Genitourinary Pain Index (GUPI) to assess pain and the ICSI to separately assess urinary symptoms. This suggestion is based on their finding of variable effects of urologic pain versus urinary symptoms on quality of life and mental health. Although symptom scores should not be relied on as diagnostic tools, their utility in establishing objective baseline measures to monitor response to treatment and symptoms over time can be valuable to the patient and the practitioner.

URINE STUDIES

Urine studies (urinalysis, culture, sensitivity, and cytology) should be included in the workup of a patient in whom IC/BPS is suspected. However, their role is mostly in ruling out other confusable disease rather than in aiding in the diagnosis of IC/BPS. A microscopic examination of the urine can reveal abnormalities attributable to the kidney that may warrant referral to a nephrologist; microscopic hematuria may trigger cystoscopic examination and referral to a urologist. The presenting symptoms of IC/BPS often mimic those of UTI, which must be ruled out by urine cultures. It is important to recognize that IC/BPS patients are subject to at least as great a risk of UTIs as the general population and that UTI should be considered when a flare in symptoms is reported. Finally, urine cytology should be considered if a diagnosis of bladder cancer is suspected or if there is a history of hematuria.

IMAGING, CYSTOSCOPY, AND URODYNAMICS

More intensive investigations and imaging studies should be considered in specific scenarios but need not be routinely performed. Abdominal and pelvic imaging studies in selected patients can help identify anatomic abnormalities of the upper or lower urinary tracts, diagnose urolithiasis or masses in the upper urinary tract, and rule out hydronephrosis, which may suggest obstructive uropathy. Furthermore, brain imaging with functional MRI and quantitative sensory testing may prove beneficial in establishing a central sensitization phenotype; however, this is an emerging field of investigation, and routine brain imaging and sensory testing are currently not recommended.

Cystoscopy is used to rule out bladder pathology—most importantly, bladder cancer. Moreover, cystoscopy plays an important role in phenotyping IC/BPS and is required for identification of Hunner lesions. Although a broad consensus is lacking, the authors and others advocate for routine cystoscopic evaluation when IC/BPS is suspected, given the potential therapeutic implications and the ability of this measure to make phenotype-directed therapies possible. Finally, urodynamics testing should be reserved for specific scenarios—for example, cases in which complex voiding dysfunction may be contributing to the presentation.

INTRAVESICAL ANESTHETIC BLADDER CHALLENGE AND HYDRODISTENSION

An intravesical anesthetic bladder challenge (using intravesical lidocaine) can be done in the outpatient setting and can help distinguish bladder-focused pain from pelvic pain of other causes. It can further be harnessed as a therapeutic strategy if the patient experiences an improvement in symptoms. Similarly, hydrodistension, which requires a general or regional anesthetic, can play a diagnostic or therapeutic role. Bladder capacities of <400 mL under general anesthesia have correlated with worse pain and poor prognosis. The diagnostic role of post-hydrodistension inspection for bladder glomerulations has been suggested as a possible important clinical differentiation, although the utility of identifying and grading glomerulations is debated.

TREATMENT

Clinical Phenotyping

The UPOINT phenotyping tool introduced in 2009 was the first clinical tool to recognize that patients presenting with pelvic pain syndromes are a heterogeneous population with disease of unclear etiology that makes it difficult to predict outcomes in individuals with standard therapies. UPOINT is based on a patient-centered approach: individualized treatments are matched to patient evaluations by phenotyping of patients using six distinct clinical domains—*u*rinary, *p*sychosocial, *o*rgan-specific, *i*nfectious, *n*eurologic, and *t*enderness. Since its initial publication, follow-up phenotyping studies have indicated that UPOINT is likely better than other methods in establishing phenotypic pain patterns in the clinic setting as local (bladder specific or pelvic pain only) or widespread (pelvic pain and beyond). Similarly, identifying inflammatory subtypes (e.g., Hunner lesion patients) and psychological parameters can help organize a patient's management plan and make it more likely that interventions will be successful. Applying an individualized multimodal treatment approach has proven beneficial in clinical practice. A collection of treatment options that might be directed at different domains of disease is presented in **Fig. 51-1**.

Although many of these treatments would be considered outside the scope of a general practitioner or general internist, it is important for the practitioner to be aware of them. In general, treatment should begin with more conservative measures, moving on to oral regimens or more invasive a multimodal approach based on the individual patient's presenting clinical phenotype or "clinical picture." Physicians are better positioned to implement this approach than are surgeons (urologists and gynecologists), who tend to be more organ- and surgery-focused when treating IC/BPS

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TREATMENT

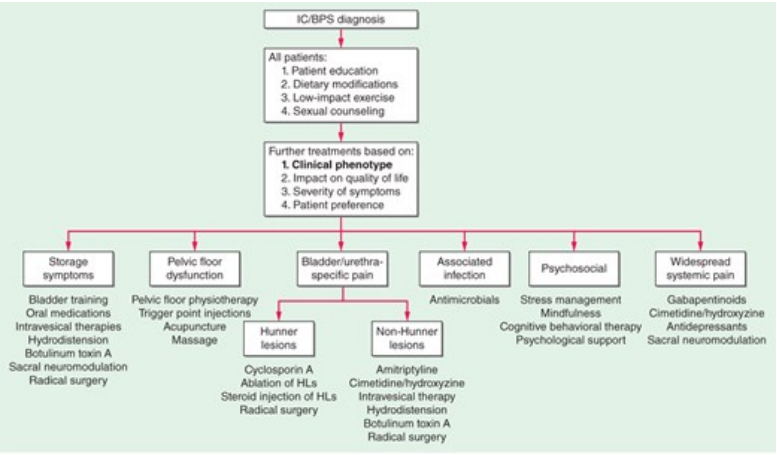
Clinical Phenotyping

The UPOINT phenotyping tool introduced in 2009 was the first clinical tool to recognize that patients presenting with pelvic pain syndromes are a heterogeneous population with disease of unclear etiology that makes it difficult to predict outcomes in individuals with standard therapies. UPOINT is based on a patient-centered approach: individualized treatments are matched to patient evaluations by phenotyping of patients using six distinct clinical domains—*u*rinary, *p*sychosocial, *o*rgan-specific, *i*nfectious, *n*eurologic, and *t*enderness. Since its initial publication, follow-up phenotyping studies have indicated that UPOINT is likely better than other methods in establishing phenotypic pain patterns in the clinic setting as local (bladder specific or pelvic pain only) or widespread (pelvic pain and beyond). Similarly, identifying inflammatory subtypes (e.g., Hunner lesion patients) and psychological parameters can help organize a patient's management plan and make it more likely that interventions will be successful. Applying an individualized multimodal treatment approach has proven beneficial in clinical practice. A collection of treatment options that might be directed at different domains of disease is presented in **Fig. 51-1**.

Although many of these treatments would be considered outside the scope of a general practitioner or general internist, it is important for the practitioner to be aware of them. In general, treatment should begin with more conservative measures, moving on to oral regimens or more invasive procedures if the patient's condition does not improve. A patient-centered approach is paramount in considering treatment escalation. The American Urological Association's IC/BPS guidelines provide a measure of overall efficacy of each individual therapy and a suggested order of implementation (tiered approach), but, because of the inability to predict the response to specific therapies, it is more clinically pragmatic to choose a multimodal approach based on the individual patient's presenting clinical phenotype or "clinical picture." Physicians are better positioned to implement this approach than are surgeons (urologists and gynecologists), who tend to be more organ- and surgery-focused when treating IC/BPS patients.

FIGURE 51-1

Proposed management paradigm for the treatment of interstitial cystitis (IC)/bladder pain syndrome (BPS). HLs, Hunner lesions.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

CONSERVATIVE MEASURES

Conservative measures should be implemented for all patients with a diagnosis of IC/BPS. These therapies tend to be simple and inexpensive to introduce, pose little risk of significant side effects, and can be intensified or abandoned on the basis of the patient's response.

Patient Education

Patient education and empowerment are paramount in this chronic pain disorder. Patients have often seen multiple practitioners prior to their diagnosis of IC/BPS. Acknowledging their suffering while educating them about their disease can go a long way in terms of relieving stress and anxiety

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome, R. Christopher Doiron; J. Curtis Nickel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Dietary Modifications

diagnosis of IC/BPS. Acknowledging their suffering while educating them about their disease can go a long way in terms of relieving stress and anxiety related to an unknown and poorly understood problem. This acknowledgement also helps to develop a therapeutic patient–provider relationship. Setting realistic expectations and understanding that cure is not the goal constitute an important first step. Several resources are available for patients to explore at their own leisure.

Dietary Modifications

Although limited evidence supports the role of dietary modifications, it has long been recognized that certain foods can trigger flares in IC/BPS patients and that simple dietary modifications can result in meaningful improvements in symptoms. Common dietary triggers include acidic and spicy foods and/or drinks, caffeinated or alcoholic beverages, artificial sweeteners, and/or gluten products; this list is by no means exhaustive, and dietary modifications should be made on an individual basis.

Pelvic Floor Physiotherapy

Involvement of the pelvic floor in the pain syndrome can be ascertained on physical examination. Randomized studies have shown that, for patients who are found to have dysfunctional pelvic floors—muscle spasm, trigger points, or tenderness—contributing to their pain syndrome, pelvic floor physiotherapy may be beneficial. The musculoskeletal anatomy of the pelvic floor is complex; finding a provider with training specifically on the pelvic floor can be difficult but is crucial. Because accessing this resource may be financially burdensome for the patient, working together to find a way to obtain this helpful adjunctive therapy is important.

Psychological Interventions

Mental health and psychosocial factors have long been identified as significantly prevalent in the IC/BPS population and can impact disease and quality-of-life outcomes. There is some indication that, in IC/BPS and other related chronic pain conditions, mindfulness and cognitive behavioral therapy may improve outcomes. Challenges in accessing these therapies are a major barrier, and there is a general lack of consensus on which specific interventions are best suited to individual patients.

MEDICAL THERAPIES

Only two medications are currently approved by the U.S. Food and Drug Administration (FDA) for the treatment of IC/BPS: pentosan polysulfate sodium (PPS) given orally and dimethyl sulfoxide (DMSO) given intravesically. However, a collection of medications, administered orally or intravesically, are commonly used (albeit off-label) for this purpose.

Oral Therapies

PPS

The only FDA-approved oral medication for IC/BPS has recently come under scrutiny because of reports regarding its association with vision-threatening maculopathy. Although causation has yet to be established, given its marginal benefit in the treatment of IC/BPS, the authors recommend against the long-term use of this medication. For patients currently taking PPS, the risks and benefits of treatment must be weighed. Consideration of a trial of weaning off the medication may be in the best interest of the patient. Any patients experiencing vision-related complaints while taking PPS should undergo immediate ophthalmologic assessment.

ANTIBIOTICS

IC/BPS is not an infectious condition, and thus, antibiotics should have no role in treatment. Furthermore, the overwhelming majority of IC/BPS patients will have received at least one course, if not several courses, of antibiotics at some point in the course of their disease. Nevertheless, it is not unreasonable to administer a single course of antibiotics (after obtaining a sample for urine culture and sensitivity testing) if the patient has never previously received such therapy.

AMITRIPTYLINE

Amitriptyline's pharmacologic activity is attributable primarily to its anticholinergic properties, its serotonin and norepinephrine uptake–inhibiting activity, and its sedative effects, which may include an antihistaminic pathway. Amitriptyline has been used to treat IC/BPS and other chronic pain

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

CIMETIDINE AND HYDROXYZINE

Early nonrandomized studies of hydroxyzine in the treatment of IC/BPS yielded promising results. As with amitriptyline, hydroxyzine's mechanism of

activity, and its sedative effects, which may include an antihistaminic pathway. Amitriptyline has been used to treat IC/BPS and other chronic pain syndromes. Studies support the use of amitriptyline in IC/BPS patients while recognizing that the benefits can be marginal and associated with significant side effects.

CIMETIDINE AND HYDROXYZINE

Early nonrandomized studies of hydroxyzine in the treatment of IC/BPS yielded promising results. As with amitriptyline, hydroxyzine's mechanism of action in treating IC/BPS is not fully understood and is likely multifactorial, owing largely to its antihistaminic effect via H_1-receptor antagonism but perhaps also its anticholinergic properties, its anxiolytic and sedative effects, and its inhibition of mast cell secretion and activation. An underpowered randomized study found no significant difference in symptom improvement between hydroxyzine and placebo.

After the modest success reported for hydroxyzine, cimetidine—an H_2-receptor antagonist—was investigated as another possible IC/BPS treatment. Only two observational studies and a single, small randomized clinical trial were completed and showed improvement in suprapubic pain and LUTS, particularly urinary frequency.

GABAPENTINOIDS

Although no randomized studies of gabapentinoids have been performed in the IC/BPS population, these agents have been shown to improve symptoms in related chronic pain conditions. Furthermore, observational studies have shown some efficacy in IC/BPS. In properly selected patients in whom neuropathic pain is suspected, this medication class may have some success.

CYCLOSPORIN A

Despite its significant side effect profile, cyclosporin A has been investigated as treatment for IC/BPS refractory to other, more standard therapies. Because of its potent anti-inflammatory properties (it is used extensively in organ transplant recipients), this drug is particularly effective in IC/BPS patients with Hunner lesions, although the improvement in symptoms is modest. Side effects, including hypertension and nephrotoxicity, must be carefully monitored for, and the medication is typically reserved for patients in whom standard therapies have failed.

Intravesical Therapies

Intravesical instillations remain a mainstay in the management of bladder-specific pain. Although this treatment modality typically requires an office visit, able and motivated patients can be trained to administer the medication at home. Patients' responses are variable, and the treatment is not curative, but it can significantly change the trajectory of disease in some patients and can rescue those experiencing symptom flares. Intravesical instillations can be administered as induction therapy; maintenance strategies have been proposed and can be effective for properly selected patients. A plethora of agents—most of them used off-label for this indication—have been investigated. The best-studied options are reviewed here.

DMSO

DMSO, a solvent with anti-inflammatory properties, has been used in intravesical treatment for IC/BPS for several years. Despite a lack of high-quality evidence (with efficacy documented in only one placebo-controlled randomized clinical trial), DMSO remains the only FDA-approved intravesical medication for IC/BPS treatment. Its use has fallen out of favor, however, largely because of its unpleasant side effect of halitosis (its elimination via the lungs is associated with a garlic-like odor). Although the degree of improvement in symptoms is highly variable (60–95%), DMSO remains in the armamentarium of intravesical therapies.

HEPARIN

Heparin, a glycosaminoglycan, was first investigated as a treatment for IC/BPS in light of the glycosaminoglycan layer deficiency theory of IC/BPS etiology; in animal models, heparin was shown to restore areas of damaged urothelium. Although there are no randomized clinical trials showing its efficacy, several observational studies have suggested benefit. Furthermore, in current practice, heparin is commonly administered with other medications as part of an intravesical "cocktail." Systemic absorption is minimal and appears not to affect coagulation parameters.

LIDOCAINE

Lidocaine is commonly used as a local anesthetic and has been investigated as an option for intravesical treatment for IC/BPS. This agent works by

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

HYALURONIC ACID AND CHONDROITIN SULFATE

Hyaluronic acid and, more recently, chondroitin sulfate have been targeted as potential intravesical therapies because of their potentially restorative

Lidocaine is commonly used as a local anesthetic and has been investigated as an option for intravesical treatment for IC/BPS. This agent works by blocking sensory nerves in the urothelium. Its absorption and efficacy increase by alkalinization, commonly through coadministration with sodium bicarbonate.

HYALURONIC ACID AND CHONDROITIN SULFATE

Hyaluronic acid and, more recently, chondroitin sulfate have been targeted as potential intravesical therapies because of their potentially restorative impact on the glycosaminoglycan layer. As is the case with most intravesical therapies, the quality of evidence is low, but there appears to be a modest benefit, with few side effects. Thus, these agents remain as options for patients whose disease is refractory to more standard therapies.

Trigger Point Injection

Injection of a local anesthetic into myofascial trigger points in the pelvic floor (identified on physical examination) is a minimally invasive, practical therapy that can be administered during an office visit and can provide relief in properly selected patients. As with all therapies for this chronic pain condition, patient selection is paramount. The evidence supporting this treatment is largely anecdotal and based on expert opinion. A small nonblinded observational trial found a 72% success rate among women diagnosed with chronic pelvic pain and trigger points on physical examination; 33% of women were completely pain free after the injection. Although robust prospective trials are needed, this modality adds to the clinician's armamentarium in the treatment of IC/BPS.

SURGICAL THERAPIES

Treatment of Hunner Lesions

A unique subset of IC/BPS patients have Hunner lesions. Direct treatment of these lesions through ablation with cauterization or laser treatment or, alternatively, injection of the lesion with a glucocorticoid improves symptoms in 70–90% of these patients. Hunner lesions tend to be recurrent, however; thus, treated patients still require follow-up, with consideration of a multimodal approach to their disease.

Hydrodistension

Recent reports confirm that one of the oldest therapies for IC/BPS, hydrodistension under general anesthesia, provides some benefit in up to 54% of patients. Short- and long-term adverse effects (including bladder perforation and long-term bladder wall fibrosis) and the temporary nature of the benefit (with symptoms typically recurring within 3–12 months) mean that repeated bladder distension may not be an ideal long-term management strategy.

Onabotulinum Toxin A

Onabotulinum toxin A (i.e., Botox) has been used to treat IC/BPS. There has, however, never been a randomized, placebo-controlled study evaluating onabotulinum toxin A injection into the detrusor muscle as monotherapy in this disease. Several randomized studies have evaluated this agent's efficacy and have shown improvements in symptoms; unfortunately, these trials often use onabotulinum toxin A in combination with hydrodistension, lack a placebo arm, and do not control for LUTS. Attributing benefit to this treatment is thus challenging. Furthermore, the side effect of acute urinary retention can be catastrophic in IC/BPS patients, whose pain is often secondary to bladder filling.

Sacral Neuromodulation

In properly selected patients, sacral neuromodulation (SNM) can provide symptom relief in IC/BPS. Its use for this purpose is off-label, and it is not FDA approved for pain therapy. Nevertheless, SNM is FDA approved for treatment of bladder overactivity—a common symptom in IC/BPS patients—that is refractory to standard therapies. Although studies evaluating improvements in pain have shown variable results, a recent meta-analysis examining 17 observational trials (but no randomized clinical trials) of the use of SNM for IC/BPS does support its efficacy, with a statistically significant pooled treatment success rate of up to 84%. Side effects related to SNM must be considered, and the procedure carries with it a high rate of revision surgery, which is needed in up to half of patients undergoing this treatment.

Radical Surgery

Radical surgery for the treatment of IC/BPS is reserved as a modality of desperation for the most refractory patients. Options range from substitution

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

COMPLICATIONS AND PROGNOSIS

Radical surgery for the treatment of IC/BPS is reserved as a modality of desperation for the most refractory patients. Options range from substitution cystoplasty to cystectomy with urinary diversion. Although improved symptoms and quality of life may result, such surgery is a potentially morbid operation and patient selection must be very specific.

COMPLICATIONS AND PROGNOSIS

IC/BPS is not unlike other chronic pain conditions in that, although a clear link has not been established with higher mortality, this condition is certainly associated with significant disability, decreased quality of life, and significant mental health morbidity. The economic impact of the disability associated with IC/BPS is similar to that of fibromyalgia, low back pain, rheumatoid arthritis, and peripheral neuropathy. Suicidal ideation is a reality in this patient population, with a reported prevalence as high as 11–23%.

For most patients, IC/BPS onset is subacute, with continuous development of the classic symptom complex over a short time and rapid (within 5 years) progression to its final stage. Symptoms then continue to wax and wane without significant overall change in symptomatology for the majority of patients. However, spontaneous improvement and/or resolution occurs in some patients, while a small subset experience subsequent deterioration to a small-capacity, fibrotic, noncompliant bladder (“end-stage bladder”). A multimodal approach to therapy, interdisciplinary involvement in patient care, particular attention to psychosocial parameters, and check-ins on mental health are important aspects of ongoing care.

GLOBAL CONSIDERATIONS

Significant challenges have been encountered in confirming the prevalence of IC/BPS, particularly globally. Prevalence estimates have ranged widely from as low as 3.5 per 100,000 women in a study of a Japanese population to as high as 20,000 per 100,000 in a self-report questionnaire study of a U.S. population. Despite these challenges, it has been recognized that IC/BPS is not simply a disease of the global West. Although robust epidemiologic studies outside North America, Europe, and some regions of Asia are lacking, it is presumed that this disease occurs at similar rates globally. This presumption may be extrapolated from epidemiologic studies of a related population and its male counterpart: CP/CPPS. These studies have shown rates of CP/CPPS in African and Asian populations that are similar to rates in North American populations.

There is no evidence to suggest that IC/BPS is phenotypically distinct in various geographic regions. Thus, this condition should be diagnosed and treated in the same ways globally. Given that its diagnosis of exclusion is based largely on history and physical examination and its treatment is based on a minimally invasive algorithm, with the focus on the patient’s clinical phenotype and the initial implementation of conservative therapeutic measures, IC/BPS can be well managed even in resource-poor settings. As with many poorly understood and difficult-to-treat conditions, the greatest barrier to its diagnosis and treatment may perhaps be its recognition.

FURTHER READING

Clemens JQ et al: Urologic chronic pelvic pain syndrome: Insights from the MAPP Research Network. Nat Rev Urol 16:187, 2019. [PubMed: 30560936]

Cox A et al: CUA guideline: Diagnosis and treatment of interstitial cystitis/bladder pain syndrome. Can Urol Assoc J 10:E136, 2016. [PubMed: 27790294]

Hanno PM et al: AUA guideline for the diagnosis and treatment of interstitial cystitis/bladder pain syndrome. J Urol 185:2162, 2011. [PubMed: 21497847]

Hanno P et al: Incontinence, in *International Consultation on Incontinence, September 2016,* vol 2, 6th ed, P Abrams et al (eds). Tokyo, ICUD ICS, 2017, pp 2203–2301.

van de Merwe JP et al: Diagnostic criteria, classification, and nomenclature for painful bladder syndrome/interstitial cystitis: An ESSIC proposal. Eur Urol 53:60, 2008. [PubMed: 17900797]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 51: Interstitial Cystitis/Bladder Pain Syndrome, R. Christopher Doiron; J. Curtis Nickel
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 52: Azotemia and Urinary Abnormalities

David B. Mount

INTRODUCTION

Normal kidney functions occur through numerous cellular processes to maintain body homeostasis. Disturbances in any of these functions can lead to abnormalities that may be detrimental to survival. Clinical manifestations of these disorders depend on the pathophysiology of renal injury and often are identified as a complex of symptoms, abnormal physical findings, and laboratory changes that constitute specific syndromes. These renal syndromes **(Table 52-1)** may arise from systemic illness or as primary renal disease. Nephrologic syndromes usually consist of several elements that reflect the underlying pathologic processes, typically including one or more of the following: (1) reduction in glomerular filtration rate (GFR), (2) abnormalities of urine sediment (red blood cells [RBCs], white blood cells [WBCs], casts, and crystals), (3) abnormal urinary excretion of serum proteins (proteinuria), (4) disturbances in urine volume (oliguria, anuria, polyuria), (5) presence of hypertension and/or expanded total body fluid volume (edema), (6) electrolyte abnormalities, and (7) in some syndromes, fever/pain. The specific combination of these findings should permit identification of one of the major nephrologic syndromes (Table 52-1) and allow differential diagnoses to be narrowed so that the appropriate diagnostic and therapeutic course can be determined. All these syndromes and their associated diseases are discussed in more detail in subsequent chapters. This chapter focuses on several aspects of renal abnormalities that are critically important for distinguishing among those processes: (1) reduction in GFR, (2) alterations of the urinary sediment and/or protein excretion, and (3) abnormalities of urinary volume.

TABLE 52-1

Initial Clinical and Laboratory Database for Defining Major Syndromes in Nephrology

SYNDROME	IMPORTANT CLUES TO DIAGNOSIS	COMMON FINDINGS	CHAP(S). DISCUSSING DISEASE-CAUSING SYNDROME
Acute or rapidly progressive renal failure	Anuria	Hypertension, hematuria	**310**, **314**, **316**, **319**
	Oliguria	Proteinuria, pyuria	
	Documented recent decline in GFR	Casts, edema	
Acute nephritis	Hematuria, RBC casts	Proteinuria	**314**
	Azotemia, reduced GFR, oliguria	Pyuria	
	Edema, hypertension	Circulatory congestion	
Chronic renal failure	Azotemia for >3 months	Proteinuria, casts	**311**
	Symptoms or signs of uremia, (late manifestation), casts	Hypocalcemia, hyperphosphatemia, hyperparathyroidism	
	Symptoms or signs of renal osteodystrophy	Polyuria, nocturia	
	Kidneys reduced in size bilaterally	Edema, hypertension	
Nephrotic syndrome	Proteinuria, with >3.5 g/24 h per 1.73 m^2	Casts	**314**

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

	Kidneys reduced in size bilaterally	Edema, hypertension	
	Broad casts in urinary sediment	Hyperkalemia, metabolic acidosis	
Nephrotic syndrome	Proteinuria, with >3.5 g/24 h per 1.73 m2	Casts	314
	Hypoalbuminemia	Lipiduria	
	Edema	Hypercoagulable state	
	Hyperlipidemia		
Asymptomatic urinary abnormalities	Hematuria		314
	Proteinuria (below nephrotic range)		
	Sterile pyuria, casts		
Urinary tract infection/pyelonephritis	Bacteriuria, with >105 cfu/mL	Hematuria	135
	Other infectious agent documented in urine	Mild azotemia and reduced GFR	
	Pyuria, leukocyte casts	Mild proteinuria	
	Frequency, urgency	Fever	
	Bladder tenderness, flank tenderness		
Renal tubular defects	Electrolyte disorders	Hematuria	315, 316
	Polyuria, nocturia	“Tubular” proteinuria (<1 g/24 h)	
	Renal calcification	Enuresis	
	Large kidneys	Electrolyte and/or acid-base abnormalities	
	Renal transport defects	Other electrolyte issues, e.g., hypomagnesemia	
Hypertension	Systolic/diastolic hypertension	Proteinuria	277, 317
		Casts	
		Azotemia	
Nephrolithiasis	Previous history of stone passage or removal	Hematuria	318
	Previous history of stone seen by x-	Pyuria	

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Urinary tract obstruction	Azotemia, oliguria, anuria	Hematuria	319

Access Provided by:
SILVERCHAIR

	Previous history of stone seen by x-ray	Pyuria	
	Renal colic	Frequency, urgency	
Urinary tract obstruction	Azotemia, oliguria, anuria	Hematuria	319
	Polyuria, nocturia, urinary retention	Pyuria	
	Slowing of urinary stream	Enuresis, dysuria	
	Large prostate, large kidneys		
	Flank tenderness, full bladder after voiding		

Abbreviations: cfu, colony-forming units; GFR; glomerular filtration rate; RBC, red blood cell.

AZOTEMIA

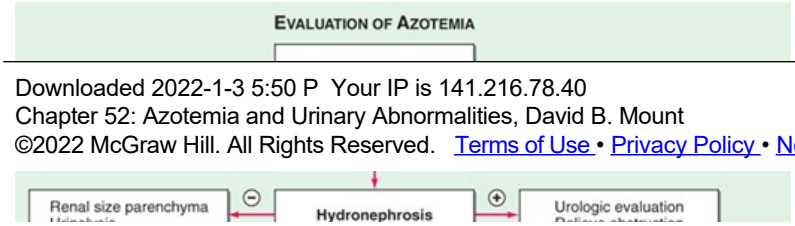
ASSESSMENT OF GFR

Monitoring the GFR is important in both hospital and outpatient settings, and several different methodologies are available. GFR is the primary metric for kidney "function," and its direct measurement involves administration of a radioactive isotope (such as inulin or iothalamate) that is filtered at the glomerulus into the urinary space but is neither reabsorbed nor secreted throughout the tubule. GFR—i.e., the clearance of inulin or iothalamate in milliliters per minute—is calculated from the rate of appearance of the isotope in the urine over several hours. In most clinical circumstances, direct GFR measurement is not feasible, and the plasma creatinine level is used as a surrogate to estimate GFR. Plasma creatinine (P_{Cr}) is the most widely used marker for GFR, which is related directly to urine creatinine (U_{Cr}) excretion and inversely to P_{Cr}. On the basis of this relationship (with some important caveats, as discussed below), GFR will fall in roughly inverse proportion to the rise in P_{Cr}. Failure to account for GFR reductions in drug dosing can lead to significant morbidity and death from drug toxicities (e.g., digoxin, imipenem). In the outpatient setting, P_{Cr} serves as an estimate for GFR (although much less accurate; see below). In patients with chronic progressive renal disease, there is an approximately linear relationship between $1/P_{Cr}$ (*y* axis) and time (*x* axis). The slope of that line will remain constant for an individual; when values deviate, an investigation for a superimposed acute process (e.g., volume depletion, drug reaction) should be initiated. Signs and symptoms of uremia, the clinical symptom complex associated with renal failure, develop at significantly different levels of P_{Cr}, depending on the patient (size, age, and sex), underlying renal disease, existence of concurrent diseases, and true GFR. Generally, patients do not develop symptomatic uremia until renal insufficiency is severe (GFR <15 mL/min).

A significantly reduced GFR (either acute or chronic) is usually reflected in a rise in P_{Cr}, leading to retention of nitrogenous waste products (defined as azotemia) such as urea. Azotemia may result from reduced renal perfusion, intrinsic renal disease, or postrenal processes (ureteral obstruction; see below and **Fig. 52-1**). Precise determination of GFR is problematic, as both commonly measured indices (urea and creatinine) have characteristics that affect their accuracy as markers of clearance. Urea clearance may underestimate GFR significantly because of urea reabsorption by the tubule. In contrast, creatinine is derived from muscle metabolism of creatine, and its generation varies little from day to day.

FIGURE 52-1

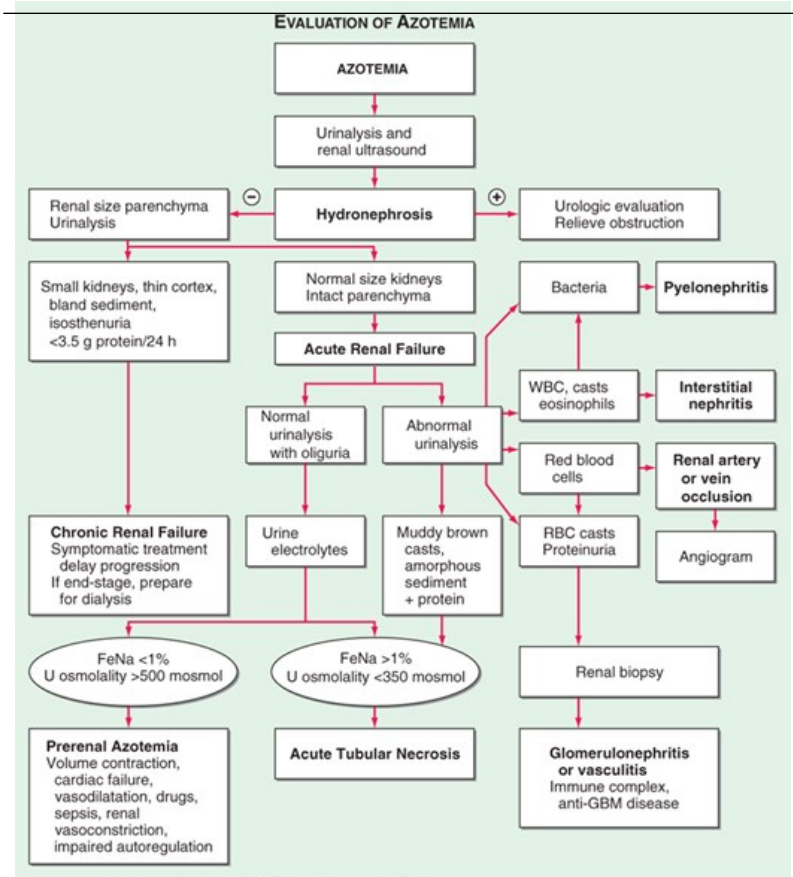
Approach to the patient with azotemia. FeNa, fractional excretion of sodium; GBM, glomerular basement membrane; RBC, red blood cell; U, urine; WBC, white blood cell.



Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Creatinine clearance (CrCl), an approximation of GFR, is measured from plasma and urinary creatinine excretion rates for a defined period (usually 24 h) and is expressed in milliliters per minute: $CrCl = (U_{vol} \times U_{Cr})/(P_{Cr} \times T_{min})$. The "adequacy" or "completeness" of the urinary collection is estimated by the urinary volume and creatinine content; creatinine is produced from muscle and excreted at a relatively constant rate. For a 20- to 50-year-old man, creatinine excretion should be 18.5–25.0 mg/kg body weight; for a woman of the same age, it should be 16.5–22.4 mg/kg body weight. For example, an 80-kg man should excrete between ~1500 and 2000 mg of creatinine in an "adequate" collection. Creatinine is useful for estimating GFR because it is a small, freely filtered solute that is not reabsorbed by the tubules. P_{Cr} levels can increase acutely from dietary ingestion of cooked meat, however, and creatinine can be secreted into the proximal tubule through an organic cation pathway (especially in advanced progressive chronic kidney disease [CKD]), leading to overestimation of GFR. When a timed collection for CrCl is not available, decisions about drug dosing must be based on P_{Cr} alone. Two formulas are used widely to estimate kidney function from P_{Cr}: (1) Cockcroft-Gault and (2) four-variable MDRD (Modification of Diet in Renal Disease).

Cockcroft-Gault:

$$\mathbf{CrCl(mL/min)} = \frac{(140 - \text{age}) \times \text{Lean Body Weight (kg)}}{\text{Serum Creatinine (mg/dL)} \times 72} (\times 0.85 \text{ if female})$$

$$\text{MDRD: eGFR (mL/min per 1.73 m}^2) = 186.3 \times P_{Cr}^{(-1.154)} \times \text{age}^{(-0.203)} \times (0.742 \text{ if female}) \times (1.21 \text{ if black}).$$

Numerous websites are available to assist with these calculations (*www.kidney.org/professionals/kdoqi/gfr_calculator.cfm*). A newer Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) estimated GFR (eGFR), which was developed by pooling several cohorts with and without kidney disease who had data on directly measured GFR, appears to be more accurate:

$$\text{CKD-EPI: eGFR} = 141 \times \min(P_{Cr}/k, 1)^{a} \times \max(P_{Cr}/k, 1)^{-1.209}$$

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

of 1, and *max* indicates the maximum of P_{Cr}/k or 1 (*https://www.mdcalc.com/ckd-epi-equations-glomerular-filtration-rate-gfr*).

$$\text{CKD-EPI: eGFR} = 141 \times \min(P_{Cr}/k, 1)^{a} \times \max(P_{Cr}/k, 1)^{-1.209} \times 0.993^{\text{Age}} \times 1.018\ [\text{if female}] \times 1.159\ [\text{if black}],$$

where P_{Cr} is plasma creatinine, *k* is 0.7 for females and 0.9 for males, *a* is −0.329 for females and −0.411 for males, *min* indicates the minimum of P_{Cr}/k or 1, and *max* indicates the maximum of P_{Cr}/k or 1 (*https://www.mdcalc.com/ckd-epi-equations-glomerular-filtration-rate-gfr*).

There are limitations to all creatinine-based estimates of GFR. Each equation, along with 24-h urine collection for measurement of creatinine clearance, is based on the assumption that the patient is in *steady state*, without daily increases or decreases in P_{Cr} as a result of rapidly changing GFR.

The MDRD equation is better correlated with true GFR when the GFR is <60 mL/min per 1.73 m^2. The gradual loss of muscle from chronic illness, chronic use of glucocorticoids, or malnutrition can mask significant changes in GFR with small or imperceptible changes in P_{Cr}.

The coefficient of 1.159 in the CKD-EPI equation to adjust for self-reported black race reflects that measured GFR was 16% higher in blacks than nonblacks with similar age, sex, and creatinine in the data set used to develop the equation. Race is a social rather than a biological construct, for which reason the use of the "race modifier" in calculating eGFR using CKD-EPI and other equations has come under scrutiny. In particular, given the implications of utilizing self-reported race to modify clinical laboratory results, many medical centers have recently stopped reporting eGFRs that have been calculated using a race modifier. This change is projected to have positive consequences, in particular, improved access to waitlisting for renal transplantation in black patients at an earlier stage of CKD. Potential negative consequences include "overdiagnosis" of CKD, inadequate or inaccurate dosing of drugs that are eliminated through the kidney (e.g., metformin), reduced access to imaging modalities for black patients with CKD with a lower reported eGFR, and reductions in living kidney donation among blacks. These and the other limitations in creatinine-based eGFR have led to the development of alternative methods for estimating GFR.

Cystatin C, a member of the cystatin superfamily of cysteine protease inhibitors, is produced at a relatively constant rate from all nucleated cells. Serum cystatin C has been proposed to be a more sensitive marker of early GFR decline than is P_{Cr}, with lesser effects of muscle mass on circulating levels; however, cystatin C levels are influenced by the patient's sex and the presence of diabetes mellitus, smoking, and inflammation. To the extent that cystatin C–based calculation of eGFR is less affected by self-reported race and muscle mass, it is an increasingly important adjunct to creatinine-based eGFR.

APPROACH TO THE PATIENT

Azotemia

Once GFR reduction has been established, the physician must decide if it represents acute or chronic renal injury. The clinical circumstances, history, and laboratory data often make this an easy distinction. However, the laboratory abnormalities characteristic of chronic renal failure, including anemia, hypocalcemia, and hyperphosphatemia, are also often present in patients presenting with acute renal failure. Radiographic evidence of renal osteodystrophy **(Chap. 311)** can be seen only in chronic renal failure but is a very late finding, typically in patients with end-stage renal disease (ESRD) maintained on dialysis. The urinalysis and renal ultrasound can facilitate distinguishing acute from chronic renal failure. An approach to the evaluation of azotemic patients is shown in Fig. 52-1. Patients with advanced chronic renal insufficiency often have some proteinuria, nonconcentrated urine (isosthenuria; isosmotic with plasma), and small kidneys on ultrasound, characterized by increased echogenicity and cortical thinning. Treatment should be directed toward slowing the progression of renal disease and providing symptomatic relief for edema, acidosis, anemia, and hyperphosphatemia, as discussed in **Chap. 311**. Acute renal failure **(Chap. 310)** can result from processes that affect blood flow and glomerular perfusion (prerenal azotemia), intrinsic renal diseases (affecting small vessels, glomeruli, or tubules), or postrenal processes (obstruction of urine flow in ureters, bladder, or urethra) **(Chap. 319)**.

PRERENAL FAILURE

Decreased renal perfusion accounts for 40–80% of cases of acute renal failure and, if appropriately treated, is readily reversible. The etiologies of prerenal azotemia include any cause of decreased circulating blood volume (gastrointestinal hemorrhage, burns, diarrhea, diuretics), volume sequestration (pancreatitis, peritonitis, rhabdomyolysis), or decreased effective arterial volume (cardiogenic shock, sepsis). Renal and glomerular perfusion also can be affected by reductions in cardiac output from peripheral vasodilation (sepsis, drugs) or profound renal vasoconstriction (severe heart failure, hepatorenal syndrome, agents such as nonsteroidal anti-inflammatory drugs [NSAIDs]). True or "effective" arterial hypovolemia leads to a fall in mean arterial pressure, which in turn triggers a series of neural and humoral responses, including activation of the sympathetic nervous and renin-angiotensin-aldosterone systems and vasopressin (AVP) release. GFR is maintained by prostaglandin-mediated

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Blockade of prostaglandin production by NSAIDs can result in severe vasoconstriction and acute renal failure. Blocking angiotensin action with angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs) decreases efferent arteriolar tone and in turn decreases

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

hypovolemia leads to a fall in mean arterial pressure, which in turn triggers a series of neural and humoral responses, including activation of the sympathetic nervous and renin-angiotensin-aldosterone systems and vasopressin (AVP) release. GFR is maintained by prostaglandin-mediated dilatation of afferent arterioles and angiotensin II–mediated constriction of efferent arterioles. Once the mean arterial pressure falls below 80 mmHg, GFR declines steeply.

Blockade of prostaglandin production by NSAIDs can result in severe vasoconstriction and acute renal failure. Blocking angiotensin action with angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs) decreases efferent arteriolar tone and in turn decreases glomerular capillary perfusion pressure. Patients taking NSAIDs and/or ACE inhibitors/ARBs are most susceptible to hemodynamically mediated acute renal failure when blood volume or arterial perfusion pressure is reduced for any reason; under these circumstances, preservation of GFR is dependent on afferent vasodilation due to prostaglandins and efferent vasoconstriction due to angiotensin II. Patients with bilateral renal artery stenosis (or stenosis in a solitary kidney) can also be dependent on efferent arteriolar vasoconstriction for maintenance of glomerular filtration pressure and are particularly susceptible to a precipitous decline in GFR when given ACE inhibitors or ARBs.

Prolonged renal hypoperfusion may lead to acute tubular necrosis (ATN), an intrinsic renal disease that is discussed below. The urinalysis and urinary electrolyte measurements can be useful in distinguishing prerenal azotemia from ATN **(Table 52-2)**. The urine Na and osmolality of patients with prerenal azotemia can be predicted from the stimulatory actions of norepinephrine, angiotensin II, AVP, aldosterone, and low tubule fluid flow rate. In prerenal conditions, the tubules are intact, leading to a concentrated urine (>500 mosmol), avid Na retention (urine Na concentration, <20 mmol/L; fractional excretion of Na [FE_{Na}], <1%), and U_{Cr}/P_{Cr} >40 (Table 52-2). The FE_{Na} is typically >1% in ATN, but may be <1% in patients with milder, nonoliguric ATN (e.g., from rhabdomyolysis) and in patients with underlying “prerenal” disorders, such as congestive heart failure (CHF) or cirrhosis or hepatorenal syndrome. The prerenal urine sediment is usually normal or has hyaline and granular casts, whereas the sediment of ATN usually is filled with cellular debris, tubular epithelial casts, and dark (muddy brown) granular casts. The measurement of urinary biomarkers associated with tubular injury is a promising technique to detect subclinical ATN and/or help further diagnose the exact cause of acute renal failure.

POSTRENAL AZOTEMIA

Urinary tract obstruction accounts for <5% of cases of acute renal failure but is usually reversible and must be ruled out early in the evaluation (Fig. 52-1). Since a single kidney is capable of adequate clearance, complete obstructive acute renal failure requires obstruction at the urethra or bladder outlet, bilateral ureteral obstruction, or unilateral obstruction in a patient with a single functioning kidney. Obstruction is usually diagnosed by the presence of ureteral and renal pelvic dilation on renal ultrasound. However, early in the course of obstruction or if the ureters are unable to dilate (e.g., encasement by pelvic or periureteral tumors or by retroperitoneal fibrosis), the ultrasound examination may be negative. Other imaging, such as a furosemide renogram (MAG3 nuclear medicine study), may be required to better define the presence or absence of obstructive uropathy. The specific urologic conditions that cause obstruction are discussed in **Chap. 319**.

INTRINSIC RENAL DISEASE

When prerenal and postrenal azotemia have been excluded as etiologies of renal failure, an intrinsic parenchymal renal disease is present. Intrinsic renal disease can arise from processes involving large renal vessels, intrarenal microvasculature and glomeruli, or the tubulointerstitium. Ischemic and toxic ATN account for ~90% of cases of acute intrinsic renal failure. As outlined in Fig. 52-1, the clinical setting and urinalysis are helpful in separating the possible etiologies. Prerenal azotemia and ATN are part of a spectrum of renal hypoperfusion; evidence of structural tubule injury is present in ATN, whereas prompt reversibility occurs with prerenal azotemia upon restoration of adequate renal perfusion. Thus, ATN often can be distinguished from prerenal azotemia by urinalysis and urine electrolyte composition (Table 52-2 and Fig. 52-1). Ischemic ATN is observed most frequently in patients who have undergone major surgery, trauma, severe hypovolemia, overwhelming sepsis, or extensive burns. Nephrotoxic ATN complicates the administration of many common medications, usually by inducing a combination of intrarenal vasoconstriction, direct tubule toxicity, and/or tubular obstruction. The kidney is vulnerable to toxic injury by virtue of its rich blood supply (25% of cardiac output) and its ability to concentrate and metabolize toxins. A diligent search for hypotension and nephrotoxins usually uncovers the specific etiology of ATN. Discontinuation of nephrotoxins and stabilization of blood pressure often suffice without the need for dialysis, with ongoing regeneration of tubular cells. **An extensive list of potential drugs and toxins implicated in ATN is found in Chap. 310.**

Processes involving the tubules and interstitium can lead to acute kidney injury (AKI), a subtype of acute renal failure. These processes include drug-induced interstitial nephritis (especially by antibiotics, NSAIDs, and diuretics), severe infections (both bacterial and viral), systemic diseases (e.g., systemic lupus erythematosus), and systemic disorders (e.g., sarcoidosis, Sjögren’s syndrome, lymphoma, or leukemia). A list of drugs associated with allergic interstitial nephritis is found in **Chap. 316**. Urinalysis usually shows mild to moderate proteinuria, hematuria, and pyuria (~75% of cases) and occasionally WBC casts. The finding of RBC casts in interstitial nephritis has been reported but should prompt a search for glomerular

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Occlusion of large renal vessels, including arteries and veins, is an uncommon cause of acute renal failure. A significant reduction in GFR by this

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

with allergic interstitial nephritis is found in **Chap. 316**. Urinalysis usually shows mild to moderate proteinuria, hematuria, and pyuria (~75% of cases) and occasionally WBC casts. The finding of RBC casts in interstitial nephritis has been reported but should prompt a search for glomerular diseases (Fig. 52-1). Occasionally, renal biopsy will be needed to distinguish among these possibilities. The classic sediment finding in allergic interstitial nephritis is a predominance (>10%) of urinary eosinophils with Wright's or Hansel's stain; however, urinary eosinophils can be increased in several other causes of AKI, such that measurement of urine eosinophils has no diagnostic utility in renal disease.

Occlusion of large renal vessels, including arteries and veins, is an uncommon cause of acute renal failure. A significant reduction in GFR by this mechanism suggests bilateral processes or, in a patient with a single functioning kidney, a unilateral process. In patients with preexisting renal artery stenosis, a substantial renal collateral circulation can develop over time and sustain renal perfusion—typically not enough to sustain glomerular filtration—in the event of total renal artery occlusion. Renal arteries can be occluded with atheroemboli, thromboemboli, in situ thrombosis, aortic dissection, or vasculitis. Atheroembolic renal failure can occur spontaneously but most often is associated with recent aortic instrumentation. The emboli are cholesterol-rich and lodge in medium and small renal arteries, with a consequent eosinophil-rich inflammatory reaction. Patients with atheroembolic acute renal failure often have a normal urinalysis, but the urine may contain eosinophils and casts. The diagnosis can be confirmed by renal biopsy, but this procedure is often unnecessary when other stigmata of atheroemboli are present (livedo reticularis, distal peripheral infarcts, eosinophilia). Renal artery thrombosis may lead to mild proteinuria and hematuria, whereas renal vein thrombosis typically occurs in the context of heavy proteinuria and hematuria. These vascular complications often require angiography for confirmation and are discussed in **Chap. 317**.

Diseases of the glomeruli (glomerulonephritis and vasculitis) and the renal microvasculature (hemolytic-uremic syndromes, thrombotic thrombocytopenic purpura, and malignant hypertension) usually present with various combinations of glomerular injury: proteinuria, hematuria, reduced GFR, and alterations of sodium excretion that lead to hypertension, edema, and circulatory congestion (acute nephritic syndrome). These findings may occur as primary renal diseases or as renal manifestations of systemic diseases. The clinical setting and other laboratory data help distinguish primary renal diseases from systemic diseases. The finding of RBC casts in the urine is an indication for early renal biopsy (Fig. 52-1), as the pathologic pattern has important implications for diagnosis, prognosis, and treatment. Hematuria without RBC casts can also be an indication of glomerular disease, since RBC casts are highly specific but very insensitive for glomerulonephritis. The specificity of urine microscopy can be enhanced by examining urine with a phase contrast microscope capable of detecting dysmorphic red cells ("acanthocytes") that are associated with glomerular disease. This evaluation is summarized in **Fig. 52-2**. A detailed discussion of glomerulonephritis and diseases of the microvasculature is found in **Chap. 316**.

OLIGURIA AND ANURIA

Oliguria refers to a 24-h urine output <400 mL, and *anuria* is the complete absence of urine formation (<100 mL). Anuria can be caused by complete bilateral urinary tract obstruction; a vascular catastrophe (dissection or arterial occlusion); renal vein thrombosis; acute cast nephropathy in myeloma; renal cortical necrosis; severe ATN; combined therapy with NSAIDs, ACE inhibitors, and/or ARBs; and hypovolemic, cardiogenic, or septic shock. Oliguria is never normal, since at least 400 mL of maximally concentrated urine must be produced to excrete the obligate daily osmolar load. *Nonoliguria* refers to urine output >400 mL/d in patients with acute or chronic azotemia. With nonoliguric ATN, disturbances of potassium and hydrogen balance are less severe than in oliguric patients, and recovery to normal renal function is usually more rapid.

TABLE 52-2

Laboratory Findings in Acute Renal Failure

INDEX	PRERENAL AZOTEMIA	OLIGURIC ACUTE RENAL FAILURE
BUN/P_{Cr} ratio	>20:1	10–15:1
Urine sodium U_{Na}, meq/L	<20	>40
Urine osmolality, mosmol/L H_2O	>500	<350
Fractional excretion of sodium[a]	<1%	>2%

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 52-2

Laboratory Findings in Acute Renal Failure

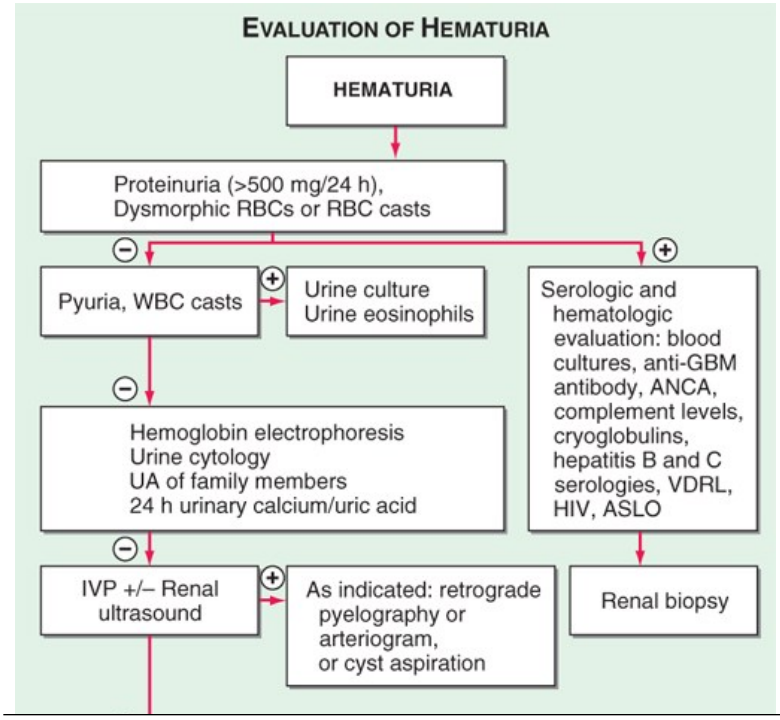
INDEX	PRERENAL AZOTEMIA	OLIGURIC ACUTE RENAL FAILURE
BUN/P_{Cr} ratio	>20:1	10–15:1
Urine sodium U_{Na}, meq/L	<20	>40
Urine osmolality, mosmol/L H_2O	>500	<350
Fractional excretion of sodium[a]	<1%	>2%
Urine/plasma creatinine U_{Cr}/P_{Cr}	>40	<20
Urinalysis (casts)	None or hyaline/granular	Muddy brown

[a] $FE_{Na} = \frac{U_{Na} \times P_{Cr} \times 100}{P_{Na} \times U_{cr}}$

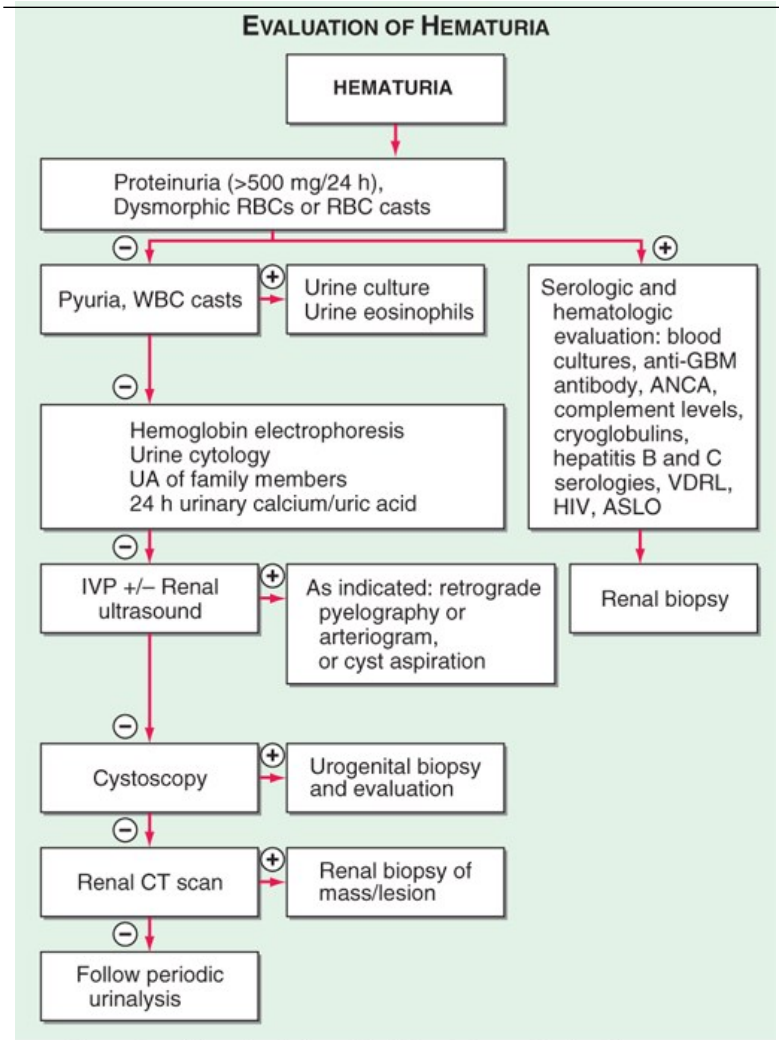
Abbreviations: BUN, blood urea nitrogen; P_{Cr}, plasma creatinine concentration; P_{Na}, plasma sodium concentration; U_{Cr}, urine creatinine concentration; U_{Na}, urine sodium concentration.

FIGURE 52-2

Approach to the patient with hematuria. ANCA, antineutrophil cytoplasmic antibody; ASLO, antistreptolysin O; CT, computed tomography; GBM, glomerular basement membrane; IVP, intravenous pyelography; RBC, red blood cell; UA, urinalysis; VDRL, Venereal Disease Research Laboratory; WBC, white blood cell.



Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

ABNORMALITIES OF THE URINE

PROTEINURIA

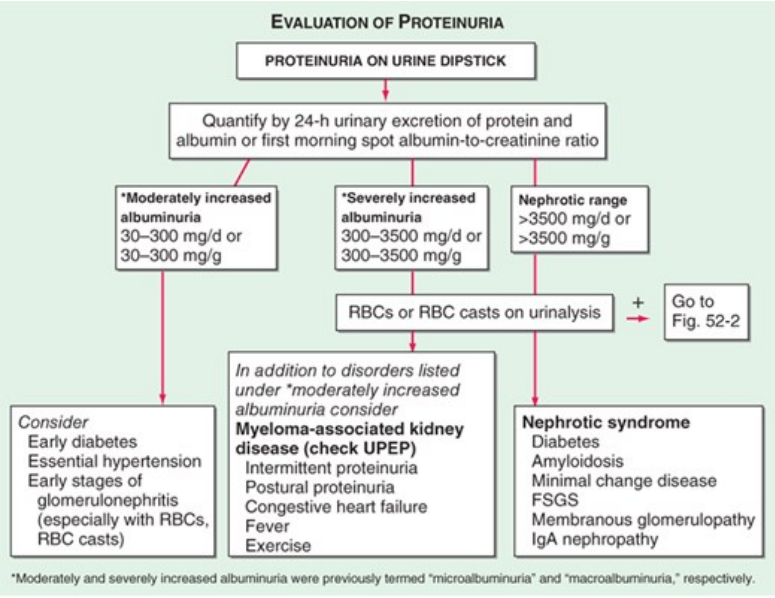
The evaluation of proteinuria is shown schematically in **Fig. 52-3** and typically is initiated after detection of proteinuria by dipstick examination. The dipstick measurement detects only albumin and gives false-positive results at pH >7.0 or when the urine is very concentrated or contaminated with blood. Because the dipstick relies on urinary albumin concentration, a very dilute urine may obscure significant proteinuria on dipstick examination. Quantification of urinary albumin on a spot urine sample (ideally from a first morning void) by measurement of an albumin-to-creatinine ratio (ACR) is helpful in approximating a 24-h albumin excretion rate (AER), where ACR (mg/g) ≈ AER (mg/24 h). Furthermore, proteinuria that is not predominantly due to albumin will be missed by dipstick screening. This information is particularly important for the detection of Bence-Jones proteins in the urine of patients with multiple myeloma. Tests to measure total urine protein concentration accurately rely on precipitation with sulfosalicylic or trichloroacetic acid (Fig. 52-3). As with albuminuria, the ratio of protein to creatinine in a random, "spot" urine can also provide a rough estimate of protein excretion; for example, a protein/creatinine ratio of 3.0 correlates to ~3.0 g of proteinuria per day. Formal assessment of urinary protein excretion requires a 24-h urine protein collection (see "Measurement of GFR," above).

FIGURE 52-3

Approach to the patient with proteinuria. Investigation of proteinuria is often initiated by a positive dipstick on routine urinalysis. Conventional urine protein electrophoresis.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Approach to the patient with proteinuria. Investigation of proteinuria is often initiated by a positive dipstick on routine urinalysis. Conventional dipsticks detect predominantly albumin and provide a semiquantitative assessment (trace, 1+, 2+, or 3+), which is influenced by urinary concentration as reflected by urine specific gravity (minimum, <1.005; maximum, 1.030). However, more exact determination of proteinuria should employ a spot morning protein/creatinine ratio (mg/g) or a 24-h urine collection (mg/24 h). FSGS, focal segmental glomerulosclerosis; RBC, red blood cell; UPEP, urine protein electrophoresis.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The magnitude of proteinuria and its composition in the urine depend on the mechanism of renal injury that leads to protein losses. Both charge and size selectivity normally prevent virtually all plasma albumin, globulins, and other high-molecular-weight proteins from crossing the glomerular wall; however, if this barrier is disrupted, plasma proteins may leak into the urine (glomerular proteinuria; Fig. 52-3). Smaller proteins (<20 kDa) are freely filtered but are readily reabsorbed by the proximal tubule. Typically, healthy individuals excrete <150 mg/d of total protein and <30 mg/d of albumin. However, even at albuminuria levels <30 mg/d, risk for progression to overt nephropathy or subsequent cardiovascular disease is increased. The remainder of the protein in the urine is secreted by the tubules (Tamm-Horsfall, IgA, and urokinase) or represents small amounts of filtered β_2-microglobulin, apoproteins, enzymes, and peptide hormones. Another mechanism of proteinuria entails excessive production of an abnormal protein that exceeds the capacity of the tubule for reabsorption. This situation most commonly occurs with plasma cell dyscrasias, such as multiple myeloma, amyloidosis, and lymphomas, that are associated with monoclonal production of immunoglobulin light chains.

The normal glomerular endothelial cell forms a barrier composed of pores of ~100 nm that retain blood cells but offer little impediment to passage of most proteins. The glomerular basement membrane traps most large proteins (>100 kDa), and the foot processes of epithelial cells (podocytes) cover the urinary side of the glomerular basement membrane and produce a series of narrow channels (slit diaphragms) to allow molecular passage of small solutes and water but not proteins. Some glomerular diseases, such as minimal change disease, cause fusion of glomerular epithelial cell foot processes, resulting in predominantly "selective" (Fig. 52-3) loss of albumin. Other glomerular diseases can present with disruption of the basement membrane and slit diaphragms (e.g., by immune complex deposition), resulting in losses of albumin and other plasma proteins. The fusion of foot processes causes increased pressure across the capillary basement membrane, resulting in areas with larger pore sizes (and more severe "nonselective" proteinuria) (Fig. 52-3).

When the total daily urinary excretion of protein is >3.5 g, hypoalbuminemia, hyperlipidemia, and edema (nephrotic syndrome; Fig. 52-3) are often present as well. However, total daily urinary protein excretion >3.5 g can occur without the other features of the nephrotic syndrome in a variety of other renal diseases, including diabetes (Fig. 52-3). Plasma cell dyscrasias (multiple myeloma) can be associated with large amounts of excreted light chains in the urine, which may not be detected by dipstick. The light chains are filtered by the glomerulus and overwhelm the reabsorptive capacity of the proximal tubule. Renal failure from these disorders occurs through a variety of mechanisms, including but not limited to proximal tubule injury, tubule obstruction (cast nephropathy), amyloid deposition, and light chain deposition **(Chap. 316)**. The specific renal lesion is dictated by the sequence and structural characteristics of the monoclonal light chain; however, not all excreted light chains are nephrotoxic.

Hypoalbuminemia in nephrotic syndrome occurs through excessive urinary losses and increased proximal tubule catabolism of filtered albumin.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

normally retained by the glomerular filtration barrier, can also directly activate sodium reabsorption by the epithelial Na channels in principal cells (ENaC) in nephrotic syndrome. Despite these changes, hypertension is uncommon in primary kidney diseases resulting in the nephrotic syndrome (Fig.

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Hypoalbuminemia in nephrotic syndrome occurs through excessive urinary losses and increased proximal tubule catabolism of filtered albumin. Edema results from renal sodium retention and reduced plasma oncotic pressure, which favors fluid movement from capillaries to interstitium. To compensate for the perceived decrease in effective intravascular volume, activation of the renin-angiotensin system, stimulation of AVP, and activation of the sympathetic nervous system take place, promoting continued renal salt and water reabsorption and progressive edema. Filtered proteases, normally retained by the glomerular filtration barrier, can also directly activate sodium reabsorption by the epithelial Na channels in principal cells (ENaC) in nephrotic syndrome. Despite these changes, hypertension is uncommon in primary kidney diseases resulting in the nephrotic syndrome (Fig. 52-3 and **Chap. 314**). The urinary loss of regulatory proteins and changes in hepatic synthesis contribute to the other manifestations of the nephrotic syndrome. A hypercoagulable state may arise from urinary losses of antithrombin III, reduced serum levels of proteins S and C, hyperfibrinogenemia, and enhanced platelet aggregation. Hypercholesterolemia may be severe and results from increased hepatic lipoprotein synthesis. Loss of immunoglobulins contributes to an increased risk of infection. Many diseases (some listed in Fig. 52-3) and drugs can cause the nephrotic syndrome; a complete list is found in **Chap. 314**.

HEMATURIA, PYURIA, AND CASTS

Isolated hematuria without proteinuria, other cells, or casts is often indicative of bleeding from the urinary tract. Hematuria is defined as two to five RBCs per high-power field (HPF) and can be detected by dipstick. A false-positive dipstick for hematuria (where no RBCs are seen on urine microscopy) may occur when myoglobinuria is present, often in the setting of rhabdomyolysis. Common causes of isolated hematuria include stones, neoplasms, tuberculosis, trauma, and prostatitis. Gross hematuria with blood clots usually is not an intrinsic renal process; rather, it suggests a postrenal source in the urinary collecting system. Evaluation of patients presenting with microscopic hematuria is outlined in Fig. 52-2. A single urinalysis with hematuria is common and can result from menstruation, viral illness, allergy, exercise, or mild trauma. Persistent or significant hematuria (>3 RBCs/HPF on three urinalyses, a single urinalysis with >100 RBCs, or gross hematuria) is associated with significant renal or urologic lesions in 9.1% of cases. The level of suspicion for urogenital neoplasms in patients with isolated painless hematuria and nondysmorphic RBCs increases with age. Neoplasms are rare in the pediatric population, and isolated hematuria is more likely to be “idiopathic” or associated with a congenital anomaly. Hematuria with pyuria and bacteriuria is typical of infection and should be treated with antibiotics after appropriate cultures. Acute cystitis or urethritis in women can cause gross hematuria. Hypercalciuria and hyperuricosuria are also risk factors for unexplained isolated hematuria in both children and adults. In some of these patients (50–60%), reducing calcium and uric acid excretion through dietary interventions can eliminate the microscopic hematuria.

Isolated microscopic hematuria can be a manifestation of glomerular diseases. The RBCs of glomerular origin are often dysmorphic when examined by phase-contrast microscopy. Irregular shapes of RBCs may also result from pH and osmolarity changes produced along the distal nephron. Observer variability in detecting dysmorphic RBCs is common. The most common etiologies of isolated glomerular hematuria are IgA nephropathy, hereditary nephritis, and thin basement membrane disease. IgA nephropathy and hereditary nephritis can lead to episodic gross hematuria. A family history of renal failure is often present in hereditary nephritis, and patients with thin basement membrane disease often have family members with microscopic hematuria. A renal biopsy is needed for the definitive diagnosis of these disorders, which are discussed in more detail in **Chap. 314**. Hematuria with dysmorphic RBCs, RBC casts, and protein excretion >500 mg/d is virtually diagnostic of glomerulonephritis. RBC casts form as RBCs that enter the tubule fluid and become trapped in a cylindrical mold of gelled Tamm-Horsfall protein. Even in the absence of azotemia, these patients should undergo serologic evaluation and renal biopsy as outlined in Fig. 52-2.

Isolated pyuria is unusual since inflammatory reactions in the kidney or collecting system also are associated with hematuria. The presence of bacteria suggests infection, and WBC casts with bacteria are indicative of pyelonephritis; “sterile pyuria” with negative urinary bacterial cultures can be seen in urogenital tuberculosis. WBCs and/or WBC casts also may be seen in acute glomerulonephritis as well as in tubulointerstitial processes such as interstitial nephritis and transplant rejection.

Casts can be seen in chronic renal diseases. Degenerated cellular casts called *waxy casts* or *broad casts* (arising in the dilated tubules that have undergone compensatory hypertrophy in response to reduced renal mass) may be seen in the urine.

ABNORMALITIES OF URINE VOLUME

POLYURIA

By history, it is often difficult for patients to distinguish urinary frequency (often of small volumes) from true polyuria (>3 L/d), and a quantification of volume by 24-h urine collection may be needed **(Fig. 52-4)**. Polyuria results from two potential mechanisms: (1) excretion of nonabsorbable solutes

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

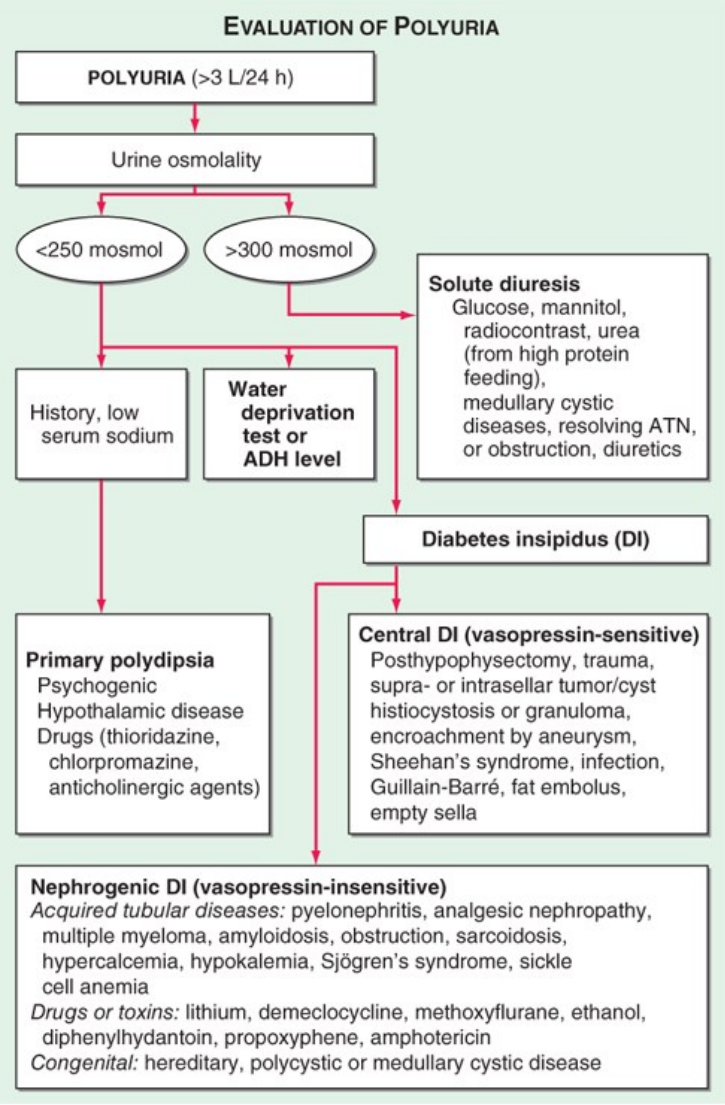
mosmol/L), total osmolar excretion is normal and a water diuresis is present. This circumstance could arise from polydipsia, inadequate secretion of AVP (*central diabetes insipidus*), or failure of renal tubules to respond to AVP (*nephrogenic diabetes insipidus*). If the urine volume is >3 L/d and urine

Access Provided by:
SILVERCHAIR

volume by 24-h urine collection may be needed **(Fig. 52-4)**. Polyuria results from two potential mechanisms: (1) excretion of nonabsorbable solutes (such as glucose) or (2) excretion of water (usually from a defect in AVP production or renal responsiveness). To distinguish a solute diuresis from a water diuresis and to determine whether the diuresis is appropriate for the clinical circumstances, urine osmolality is measured. The average person excretes between 600 and 800 mosmol of solutes per day, primarily as urea and electrolytes. If the urine output is >3 L/d and the urine is dilute (<250 mosmol/L), total osmolar excretion is normal and a water diuresis is present. This circumstance could arise from polydipsia, inadequate secretion of AVP (*central diabetes insipidus*), or failure of renal tubules to respond to AVP (*nephrogenic diabetes insipidus*). If the urine volume is >3 L/d and urine osmolality is >300 mosmol/L, a solute diuresis is clearly present and a search for the responsible solute(s) is mandatory.

FIGURE 52-4

Approach to the patient with polyuria. ADH, antidiuretic hormone; ATN, acute tubular necrosis.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Excessive filtration of a poorly reabsorbed solute such as glucose or mannitol can depress reabsorption of NaCl and water in the proximal tubule and lead to enhanced excretion in the urine. Poorly controlled diabetes mellitus with glucosuria is the most common cause of a solute diuresis, leading to volume depletion and serum hypertonicity. Since the urine sodium concentration is less than that of blood, more water than sodium is lost, causing hypernatremia and hypertonicity. Common iatrogenic solute diuresis occurs in association with mannitol administration, radiocontrast media, and high-protein feedings (enteral or parenteral), leading to increased urea production and excretion. Less commonly, excessive sodium loss may result from cystic renal diseases or Bartter's syndrome or may develop during a tubulointerstitial process (such as resolving ATN). In these so-called salt-wasting disorders, the tubule damage results in direct impairment of sodium reabsorption and indirectly reduces the responsiveness of the tubule to

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Formation of large volumes of dilute urine is usually due to polydipsic states or diabetes insipidus. Primary polydipsia can result from habit, psychiatric disorders, neurologic lesions, or medications. During deliberate polydipsia, extracellular fluid volume is normal or expanded and plasma

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

from cystic renal diseases or Bartter's syndrome or may develop during a tubulointerstitial process (such as resolving ATN). In these so-called salt-wasting disorders, the tubule damage results in direct impairment of sodium reabsorption and indirectly reduces the responsiveness of the tubule to aldosterone. Usually, the sodium losses are mild, and the obligatory urine output is <2 L/d; resolving ATN and postobstructive diuresis are exceptions and may be associated with significant natriuresis and polyuria.

Formation of large volumes of dilute urine is usually due to polydipsic states or diabetes insipidus. Primary polydipsia can result from habit, psychiatric disorders, neurologic lesions, or medications. During deliberate polydipsia, extracellular fluid volume is normal or expanded and plasma AVP levels are reduced because serum osmolality tends to be near the lower limits of normal. Urine osmolality is also maximally dilute at 50 mosmol/L.

Central diabetes insipidus may be idiopathic in origin or secondary to a variety of conditions, including hypophysectomy, trauma, neoplastic, inflammatory, vascular, or infectious hypothalamic diseases. Idiopathic central diabetes insipidus is associated with selective destruction of the AVP-secreting neurons in the supraoptic and paraventricular nuclei and can either be inherited as an autosomal dominant trait or occur spontaneously. Nephrogenic diabetes insipidus can occur in a variety of clinical situations, as summarized in Fig. 52-4.

A plasma AVP level is recommended as the best method for distinguishing between central and nephrogenic diabetes insipidus. Assays for circulating copeptin, a peptide that is cleaved from pre-pro-AVP during axonal transport in the posterior pituitary, are also now available in many centers. A water deprivation test plus exogenous AVP may distinguish primary polydipsia from central and nephrogenic diabetes insipidus. Measurement of hypertonic saline–stimulated plasma copeptin, if available, can substitute for water deprivation testing. **For a detailed discussion, see Chap. 381.**

ACKNOWLEDGMENT

Julie Lin and Brad Denker contributed to this chapter in the 19th edition, and some material from that chapter has been retained here.

FURTHER READING

Emmett M et al: Approach to the patient with kidney disease, in *Brenner and Rector's The Kidney*, 10th ed, K Skorecki et al (eds). Philadelphia, W.B. Saunders & Company, 2016, pp 754–779.

Eneanya ND et al: Reconsidering the consequences of using race to estimate kidney function. JAMA 322:113, 2019. [PubMed: 31169890]

Köhler H et al: Acanthocyturia—a characteristic marker for glomerular bleeding. Kidney Int 40:115, 1991. [PubMed: 1921146]

Perazella MA: The urine sediment as a biomarker of kidney disease. Am J Kidney Dis 66:748, 2015. [PubMed: 25943719]

Powe NR: Black kidney function matters: Use or misuse of race? JAMA 324:737, 2020. [PubMed: 32761164]

Weisord SD et al: Prevention and management of acute kidney injury in *Brenner and Rector's The Kidney*, 11th ed, ASL Yu et al: (eds). Philadelphia, W.B. Saunders & Company, 2020, pp 940–977.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 52: Azotemia and Urinary Abnormalities, David B. Mount
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

Chapter 54: Hypercalcemia and Hypocalcemia

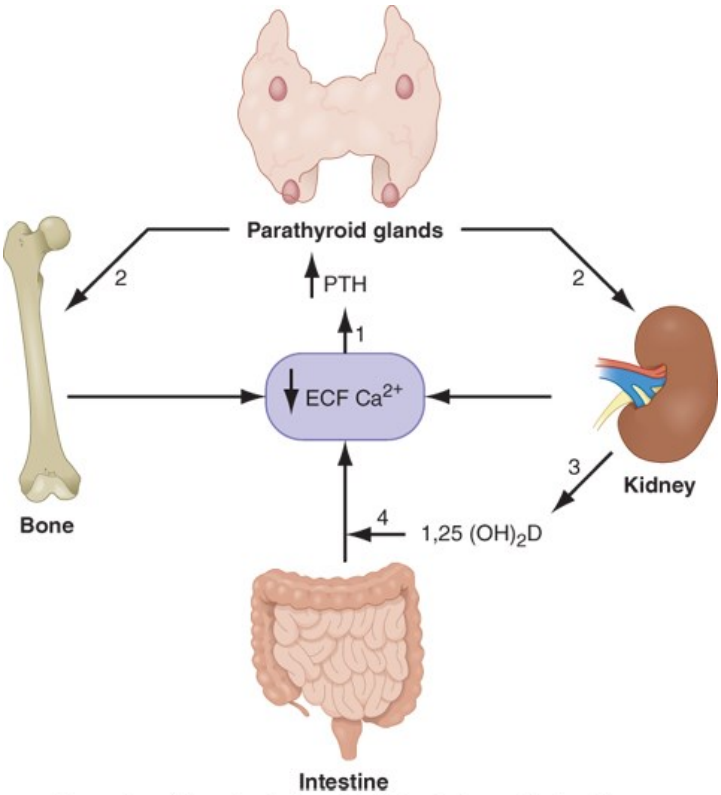
Sundeep Khosla

INTRODUCTION

The calcium ion plays a critical role in normal cellular function and signaling, regulating diverse physiologic processes such as neuromuscular signaling, cardiac contractility, hormone secretion, and blood coagulation. Thus, extracellular calcium concentrations are maintained within an exquisitely narrow range through a series of feedback mechanisms that involve parathyroid hormone (PTH) and the active vitamin D metabolite 1,25-dihydroxyvitamin D [1,25$(OH)_2$D]. These feedback mechanisms are orchestrated by integrating signals between the parathyroid glands, kidney, intestine, and bone **(Fig. 54-1; Chap. 409)**. Disorders of serum calcium concentration are relatively common and often serve as a harbinger of underlying disease. This chapter provides a brief summary of the approach to patients with altered serum calcium levels. **See Chap. 410 for a detailed discussion of this topic.**

FIGURE 54-1

Feedback mechanisms maintaining extracellular calcium concentrations within a narrow, physiologic range (8.9–10.1 mg/dL [2.2–2.5 m*M*]). A decrease in extracellular (ECF) calcium (Ca^{2+}) triggers an increase in parathyroid hormone (PTH) secretion (1) via the calcium sensor receptor on parathyroid cells. PTH, in turn, results in increased tubular reabsorption of calcium by the kidney (2) and resorption of calcium from bone (2) and also stimulates renal 1,25$(OH)_2$D production (3). 1,25$(OH)_2$D, in turn, acts principally on the intestine to increase calcium absorption (4). Collectively, these homeostatic mechanisms serve to restore serum calcium levels to normal.



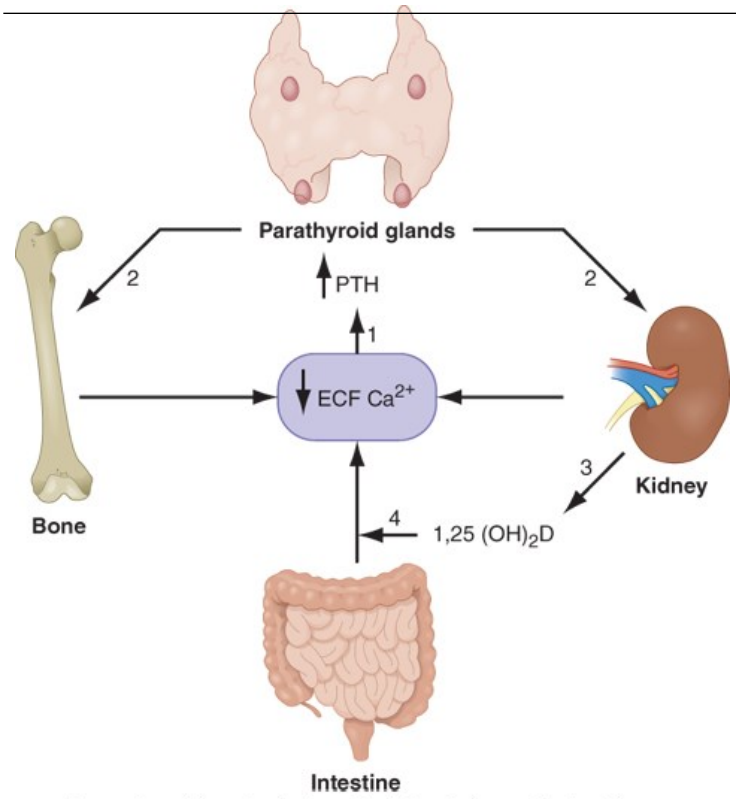
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 54: Hypercalcemia and Hypocalcemia, Sundeep Khosla
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ETIOLOGY

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

HYPERCALCEMIA

ETIOLOGY

The causes of hypercalcemia can be understood and classified based on derangements in the normal feedback mechanisms that regulate serum calcium **(Table 54-1)**. Excess PTH production, which is not appropriately suppressed by increased serum calcium concentrations, occurs in primary neoplastic disorders of the parathyroid glands (parathyroid adenomas; hyperplasia; or, rarely, carcinoma) that are associated with increased parathyroid cell mass and impaired feedback inhibition by calcium. Inappropriate PTH secretion for the ambient level of serum calcium also occurs in familial hypocalciuric hypercalcemia (FHH), which is an autosomal dominant syndrome most commonly involving inactivating mutations in the calcium sensor receptor (*CaSR*; FHH type 1), with rare families having mutations in the $G\alpha_{11}$ protein (*GNA11*; FHH type 2) or the adaptor-related protein complex 2, σ-2 subunit (*AP2S1*; FHH type 3); all of these mutations impair extracellular calcium sensing by the parathyroid glands and the kidneys, leading to inappropriate PTH secretion and increased renal tubular calcium reabsorption. Although PTH secretion by tumors is extremely rare, many solid tumors produce PTH-related peptide (PTHrP), which shares homology with PTH in the first 13 amino acids and binds the PTH receptor, thus mimicking effects of PTH on bone and the kidney. In PTHrP-mediated hypercalcemia of malignancy, PTH levels are suppressed by the high serum calcium levels. Hypercalcemia associated with granulomatous disease (e.g., sarcoidosis) or lymphomas is caused by enhanced conversion of 25(OH)D to the potent $1,25(OH)_2D$. In these disorders, $1,25(OH)_2D$ enhances intestinal calcium absorption, resulting in hypercalcemia and suppressed PTH. Disorders that directly increase calcium mobilization from bone, such as hyperthyroidism or osteolytic metastases, also lead to hypercalcemia with suppressed PTH secretion as does exogenous calcium overload, as in milk-alkali syndrome, or total parenteral nutrition with excessive calcium supplementation.

TABLE 54-1

Causes of Hypercalcemia

Excessive PTH production
Ectopic PTH secretion (very rare)

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 54: Hypercalcemia and Hypocalcemia, Sundeep Khosla
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

secretion as does exogenous calcium overload, as in milk-alkali syndrome, or total parenteral nutrition with excessive calcium supplementation.

TABLE 54-1

Causes of Hypercalcemia

Excessive PTH production
Primary hyperparathyroidism (adenoma, hyperplasia, rarely carcinoma)
Tertiary hyperparathyroidism (long-term stimulation of PTH secretion in renal insufficiency)
Ectopic PTH secretion (very rare)
FHH
Alterations in CaSR function (lithium therapy)
Hypercalcemia of malignancy
Overproduction of PTHrP (many solid tumors)
Lytic skeletal metastases (breast, myeloma)
Excessive $1,25(OH)_2D$ production
Granulomatous diseases (sarcoidosis, tuberculosis, silicosis)
Lymphomas
Vitamin D intoxication
Primary increase in bone resorption
Hyperthyroidism
Immobilization
Excessive calcium intake
Milk-alkali syndrome
Total parenteral nutrition
Other causes
Endocrine disorders (adrenal insufficiency, pheochromocytoma, VIPoma)
Medications (thiazides, vitamin A, antiestrogens)

Abbreviations: CaSR, calcium sensor receptor; FHH, familial hypocalciuric hypercalcemia; PTH, parathyroid hormone; PTHrP, PTH-related peptide.

CLINICAL MANIFESTATIONS

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 54: Hypercalcemia and Hypocalcemia, Sundeep Khosla
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

include peptic ulcer disease or nephrolithiasis, and fracture risk may be increased. More severe hypercalcemia (>12–13 mg/dL), particularly if it develops acutely, may result in lethargy, stupor, or coma, as well as gastrointestinal symptoms (nausea, anorexia, constipation, or pancreatitis).

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Mild hypercalcemia (up to 11–11.5 mg/dL) is usually asymptomatic and recognized only on routine calcium measurements. Some patients may complain of vague neuropsychiatric symptoms, including trouble concentrating, personality changes, or depression. Other presenting symptoms may include peptic ulcer disease or nephrolithiasis, and fracture risk may be increased. More severe hypercalcemia (>12–13 mg/dL), particularly if it develops acutely, may result in lethargy, stupor, or coma, as well as gastrointestinal symptoms (nausea, anorexia, constipation, or pancreatitis). Hypercalcemia decreases renal concentrating ability, which may cause polyuria and polydipsia. With long-standing hyperparathyroidism, patients may present with bone pain or pathologic fractures. Finally, hypercalcemia can result in significant electrocardiographic changes, including bradycardia, atrioventricular (AV) block, and short QT interval; changes in serum calcium can be monitored by following the QT interval.

DIAGNOSTIC APPROACH

The first step in the diagnostic evaluation of hyper- or hypocalcemia is to ensure that the alteration in serum calcium levels is not due to abnormal albumin concentrations. About 50% of total calcium is ionized, and the rest is bound principally to albumin. Although direct measurements of ionized calcium are possible, they are easily influenced by collection methods and other artifacts; thus, it is generally preferable to measure total calcium and albumin to "correct" the serum calcium. When serum albumin concentrations are reduced, a corrected calcium concentration is calculated by adding 0.2 m*M* (0.8 mg/dL) to the total calcium level for every decrement in serum albumin of 1.0 g/dL below the reference value of 4.1 g/dL for albumin, and, conversely, for elevations in serum albumin.

A detailed history may provide important clues regarding the etiology of the hypercalcemia (Table 54-1). Chronic hypercalcemia is most commonly caused by primary hyperparathyroidism, as opposed to the second most common etiology of hypercalcemia, an underlying malignancy. The history should include medication use, previous neck surgery, and systemic symptoms suggestive of sarcoidosis or lymphoma.

Once true hypercalcemia is established, the second most important laboratory test in the diagnostic evaluation is a PTH level using a two-site assay for the intact hormone. Increases in PTH are often accompanied by hypophosphatemia. In addition, serum creatinine should be measured to assess renal function; hypercalcemia may impair renal function, and renal clearance of PTH may be altered depending on the fragments detected by the assay. If the PTH level is increased (or "inappropriately normal") in the setting of elevated calcium and low phosphorus, the diagnosis is almost always primary hyperparathyroidism. Because individuals with FHH may also present with mildly elevated PTH levels and hypercalcemia, this diagnosis should be considered and excluded because parathyroid surgery is ineffective in this condition. A calcium/creatinine clearance ratio (calculated as urine calcium/serum calcium divided by urine creatinine/serum creatinine) of <0.01 is suggestive of FHH, particularly when there is a family history of mild, asymptomatic hypercalcemia. In addition, sequence analysis of the *CASR* gene is now commonly performed for the definitive diagnosis of FHH, although as noted above, in rare families, FHH may be caused by mutations in the *GNA11* or *AP2S1* genes, and patients may have to pay out-of-pocket for the genetic analysis. Ectopic PTH secretion is extremely rare.

A suppressed PTH level in the face of hypercalcemia is consistent with non-parathyroid-mediated hypercalcemia, most often due to underlying malignancy. Although a tumor that causes hypercalcemia is generally overt, a PTHrP level may be needed to establish the diagnosis of hypercalcemia of malignancy. Serum $1,25(OH)_2D$ levels are increased in granulomatous disorders, and clinical evaluation in combination with laboratory testing will generally provide a diagnosis for the various disorders listed in Table 54-1.

TREATMENT

Hypercalcemia

Mild, asymptomatic hypercalcemia does not require immediate therapy, and management should be dictated by the underlying diagnosis. By contrast, significant, symptomatic hypercalcemia usually requires therapeutic intervention independent of the etiology of hypercalcemia. Initial therapy of significant hypercalcemia begins with volume expansion because hypercalcemia invariably leads to dehydration; 4–6 L of intravenous saline may be required over the first 24 h, keeping in mind that underlying comorbidities (e.g., congestive heart failure) may require the use of loop diuretics to enhance sodium and calcium excretion. However, loop diuretics should not be initiated until the volume status has been restored to normal. If there is increased calcium mobilization from bone (as in malignancy or severe hyperparathyroidism), drugs that inhibit bone resorption should be considered. Although salmon calcitonin (4–8 IU/kg intramuscularly or subcutaneously every 6–12 h) is sometimes used, the mainstays of therapy are bisphosphonates, which are potent inhibitors of bone resorption. Zoledronic acid (e.g., 4 mg intravenously over ~30 min) and pamidronate (e.g., 60–90 mg intravenously over 2–4 h) are bisphosphonates that are commonly used for the treatment of hypercalcemia of malignancy in adults. Onset of action is within 1–3 days, with normalization of serum calcium levels occurring in 60–90% of patients refractory to bisphosphonates. An alternative to the bisphosphonates or denosumab is gallium nitrate (200 mg/m² intravenously daily for 5 days), which is also effective, but has potential nephrotoxicity. In rare instances, dialysis may be necessary. Finally, although intravenous phosphate

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 54: Hypercalcemia and Hypocalcemia, Sundeep Khosla
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TREATMENT

Hypercalcemia

Mild, asymptomatic hypercalcemia does not require immediate therapy, and management should be dictated by the underlying diagnosis. By contrast, significant, symptomatic hypercalcemia usually requires therapeutic intervention independent of the etiology of hypercalcemia. Initial therapy of significant hypercalcemia begins with volume expansion because hypercalcemia invariably leads to dehydration; 4–6 L of intravenous saline may be required over the first 24 h, keeping in mind that underlying comorbidities (e.g., congestive heart failure) may require the use of loop diuretics to enhance sodium and calcium excretion. However, loop diuretics should not be initiated until the volume status has been restored to normal. If there is increased calcium mobilization from bone (as in malignancy or severe hyperparathyroidism), drugs that inhibit bone resorption should be considered. Although salmon calcitonin (4–8 IU/kg intramuscularly or subcutaneously every 6–12 h) is sometimes used, the mainstays of therapy are bisphosphonates, which are potent inhibitors of bone resorption. Zoledronic acid (e.g., 4 mg intravenously over ~30 min) and pamidronate (e.g., 60–90 mg intravenously over 2–4 h) are bisphosphonates that are commonly used for the treatment of hypercalcemia of malignancy in adults. Onset of action is within 1–3 days, with normalization of serum calcium levels occurring in 60–90% of patients. Bisphosphonate infusions may need to be repeated if hypercalcemia relapses. Denosumab (120 mg subcutaneously on days 1, 8, 15, and 29, and then every 4 weeks), an antibody to RANKL, is a potent inhibitor of bone resorption and has been shown to be effective in treating hypercalcemia refractory to bisphosphonates. An alternative to the bisphosphonates or denosumab is gallium nitrate (200 mg/m^2 intravenously daily for 5 days), which is also effective, but has potential nephrotoxicity. In rare instances, dialysis may be necessary. Finally, although intravenous phosphate chelates calcium and decreases serum calcium levels, this therapy can be toxic because calcium-phosphate complexes may deposit in tissues and cause extensive organ damage.

In patients with 1,25$(OH)_2$D-mediated hypercalcemia, glucocorticoids are the preferred therapy, as they decrease 1,25$(OH)_2$D production. Intravenous hydrocortisone (100–300 mg daily) or oral prednisone (40–60 mg daily) for 3–7 days is used most often. Other drugs, such as ketoconazole, chloroquine, and hydroxychloroquine, may also decrease 1,25$(OH)_2$D production and are used occasionally.

HYPOCALCEMIA

ETIOLOGY

The causes of hypocalcemia can be differentiated according to whether serum PTH levels are low (hypoparathyroidism) or high (secondary hyperparathyroidism). Although there are many potential causes of hypocalcemia, impaired PTH production and impaired vitamin D production are the most common etiologies **(Table 54-2) (Chap. 410)**. Because PTH is the main defense against hypocalcemia, disorders associated with deficient PTH production or secretion may be associated with profound, life-threatening hypocalcemia. In adults, hypoparathyroidism most commonly results from inadvertent damage to all four glands during thyroid or parathyroid gland surgery. Hypoparathyroidism is a cardinal feature of autoimmune endocrinopathies **(Chap. 388)**; rarely, it may be associated with infiltrative diseases such as sarcoidosis. Impaired PTH secretion may be secondary to magnesium deficiency or to activating mutations in the CaSR or in the G proteins that mediate CaSR signaling (autosomal dominant hypocalcemia), which suppress PTH, leading to effects that are opposite to those that occur in FHH.

TABLE 54-2

Causes of Hypocalcemia

Low Parathyroid Hormone Levels (Hypoparathyroidism)
Parathyroid agenesis
Isolated
DiGeorge's syndrome
Radiation

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 54: Hypercalcemia and Hypocalcemia, Sundeep Khosla
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Parathyroid destruction
Surgical
Radiation
Infiltration by metastases or systemic diseases
Autoimmune
Reduced parathyroid function
Hypomagnesemia
Autosomal dominant hypocalcemia
High Parathyroid Hormone Levels (Secondary Hyperparathyroidism)
Vitamin D deficiency or impaired $1,25(OH)_2D$ production/action
Nutritional vitamin D deficiency (poor intake or absorption)
Renal insufficiency with impaired $1,25(OH)_2D$ production
Vitamin D resistance, including receptor defects
Parathyroid hormone resistance syndromes
PTH receptor mutations
Pseudohypoparathyroidism (G protein mutations)
Drugs
Calcium chelators
Inhibitors of bone resorption (bisphosphonates, plicamycin)
Altered vitamin D metabolism (phenytoin, ketoconazole)
Miscellaneous causes
Acute pancreatitis
Acute rhabdomyolysis
Hungry bone syndrome after parathyroidectomy
Osteoblastic metastases with marked stimulation of bone formation (prostate cancer)

Abbreviation: PTH, parathyroid hormone.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 54: Hypercalcemia and Hypocalcemia, Sundeep Khosla

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

capable of mounting a compensatory increase in PTH secretion. Hypocalcemia may also occur in conditions associated with severe tissue injury such as burns, rhabdomyolysis, tumor lysis, or pancreatitis. The cause of hypocalcemia in these settings may include a combination of low albumin,

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Abbreviation: PTH, parathyroid hormone.

Vitamin D deficiency, impaired 1,25(OH)$_2$D production (primarily secondary to renal insufficiency), or vitamin D resistance also cause hypocalcemia.

However, the degree of hypocalcemia in these disorders is generally not as severe as that seen with hypoparathyroidism because the parathyroids are capable of mounting a compensatory increase in PTH secretion. Hypocalcemia may also occur in conditions associated with severe tissue injury such as burns, rhabdomyolysis, tumor lysis, or pancreatitis. The cause of hypocalcemia in these settings may include a combination of low albumin, hyperphosphatemia, tissue deposition of calcium, and impaired PTH secretion.

CLINICAL MANIFESTATIONS

Patients with hypocalcemia may be asymptomatic if the decreases in serum calcium are relatively mild and chronic, or they may present with life-threatening complications. Moderate to severe hypocalcemia is associated with paresthesias, usually of the fingers, toes, and circumoral regions, and is caused by increased neuromuscular irritability. On physical examination, a Chvostek's sign (twitching of the circumoral muscles in response to gentle tapping of the facial nerve just anterior to the ear) may be elicited, although it is also present in ~10% of normal individuals. Carpal spasm may be induced by inflation of a blood pressure cuff to 20 mmHg above the patient's systolic blood pressure for 3 min (Trousseau's sign). Severe hypocalcemia can induce seizures, carpopedal spasm, bronchospasm, laryngospasm, and prolongation of the QT interval.

DIAGNOSTIC APPROACH

In addition to measuring serum calcium, it is useful to determine albumin, phosphorus, and magnesium levels. As for the evaluation of hypercalcemia, determining the PTH level is central to the evaluation of hypocalcemia. A suppressed (or "inappropriately low") PTH level in the setting of hypocalcemia establishes absent or reduced PTH secretion (hypoparathyroidism) as the cause of the hypocalcemia. Further history will often elicit the underlying cause (i.e., parathyroid agenesis vs destruction). By contrast, an elevated PTH level (secondary hyperparathyroidism) should direct attention to the vitamin D axis as the cause of the hypocalcemia. Nutritional vitamin D deficiency is best assessed by obtaining serum 25-hydroxyvitamin D levels, which reflect vitamin D stores. In the setting of renal insufficiency or suspected vitamin D resistance, serum 1,25(OH)$_2$D levels are informative.

TREATMENT

Hypocalcemia

The approach to treatment depends on the severity of the hypocalcemia, the rapidity with which it develops, and the accompanying complications (e.g., seizures, laryngospasm). Acute, symptomatic hypocalcemia is initially managed with calcium gluconate, 10 mL 10% wt/vol (90 mg or 2.2 mmol) intravenously, diluted in 50 mL of 5% dextrose or 0.9% sodium chloride, given intravenously over 5 min. Continuing hypocalcemia often requires a constant intravenous infusion (typically 10 ampules of calcium gluconate or 900 mg of calcium in 1 L of 5% dextrose or 0.9% sodium chloride administered over 24 h). Accompanying hypomagnesemia, if present, should be treated with appropriate magnesium supplementation.

Chronic hypocalcemia due to hypoparathyroidism is treated with calcium supplements (1000–1500 mg/d elemental calcium in divided doses) and either vitamin D$_2$ or D$_3$ (25,000–100,000 U daily) or calcitriol [1,25(OH)$_2$D, 0.25–2 μg/d]. Other vitamin D metabolites (dihydrotachysterol, alfacalcidiol) are now used less frequently. Importantly, PTH (1-84) (Natpara) is now approved by the Food and Drug Administration for the treatment of refractory hypoparathyroidism, representing an important advance in treatment of these patients. Vitamin D deficiency is best treated using vitamin D supplementation, with the dose depending on the severity of the deficit and the underlying cause. Thus, nutritional vitamin D deficiency generally responds to relatively low doses of vitamin D (50,000 IU, 2–3 times per week for several months), whereas vitamin D deficiency due to malabsorption may require much higher doses (100,000 IU/d or more). The treatment goal is to bring serum calcium into the low normal range and to avoid hypercalciuria, which may lead to nephrolithiasis.

GLOBAL CONSIDERATIONS

In countries with more limited access to health care or screening laboratory testing of serum calcium levels, primary hyperparathyroidism often presents in its severe form with skeletal complications (osteitis fibrosa cystica) in contrast to the asymptomatic form that is common in developed countries. In addition, vitamin D deficiency is paradoxically common in some countries despite extensive sunlight (e.g., India) due to avoidance of sun

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 54: Hypercalcemia and Hypocalcemia, Sundeep Khosla
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Bilezikian JP, et al: Hyperparathyroidism. Lancet 391:168, 2018. [PubMed: 28923463]

countries. In addition, vitamin D deficiency is paradoxically common in some countries despite extensive sunlight (e.g., India) due to avoidance of sun exposure and poor dietary vitamin D intake.

FURTHER READING

Bilezikian JP et al: Hyperparathyroidism. Lancet 391:168, 2018. [PubMed: 28923463]

Brandi ML et al: Management of hypoparathyroidism: Summary statement and guidelines. J Clin Endocrinol Metab 101:2273, 2016. [PubMed: 26943719]

Hannan FM et al: The calcium-sensing receptor in physiology and in calcitropic and noncalcitropic diseases. Nat Rev Endocrinol 15:33, 2018. [PubMed: 30443043]

Minisola S et al: The diagnosis and management of hypercalcemia. BMJ 350:h2723, 2015. [PubMed: 26037642]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 54: Hypercalcemia and Hypocalcemia, Sundeep Khosla
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 55: Acidosis and Alkalosis

Thomas D. DuBose, Jr.

NORMAL ACID-BASE HOMEOSTASIS

Systemic arterial pH is maintained between 7.35 and 7.45 by extracellular and intracellular chemical buffering together with respiratory and renal regulatory mechanisms. The control of arterial CO_2 tension ($Paco_2$) by the central nervous system (CNS) and respiratory system and the control of plasma bicarbonate by the kidneys stabilize the arterial pH by excretion or retention of acid or alkali. The metabolic and respiratory components that regulate systemic pH are described by the Henderson-Hasselbalch equation and solved for pH when the solubility of CO_2 is considered (dissolved CO_2 in mmol/L = 0.03 × $Paco_2$ in mmHg), at a pK' of 6.1:

$$pH = pK' + \log_{10}\frac{[HCO_3^-]}{\alpha_{CO_2} PCO_2}$$

Under most circumstances, CO_2 production and excretion are matched, and the usual steady-state $Paco_2$ is maintained at ~40 mmHg. Underexcretion of CO_2 produces hypercapnia, and overexcretion causes hypocapnia. Nevertheless, production and excretion are again matched at a new steady-state $Paco_2$. Therefore, the $Paco_2$ is regulated primarily by neural respiratory factors and is not subject to regulation by the rate of CO_2 production. Hypercapnia is usually the result of hypoventilation rather than of increased CO_2 production. Increases or decreases in $Paco_2$ represent derangements of neural respiratory control or are due to compensatory changes in response to a primary alteration in the plasma $[HCO_3^-]$.

DIAGNOSIS OF GENERAL TYPES OF DISTURBANCES

The most common clinical disturbances are simple acid-base disorders; i.e., metabolic acidosis or alkalosis or respiratory acidosis or alkalosis occurring individually. Recognition of simple acid-base disorders requires appreciation of the limits of physiologic compensation for a primary disturbance.

SIMPLE ACID-BASE DISORDERS

Primary respiratory disturbances (primary changes in $Paco_2$) invoke compensatory metabolic responses (secondary changes in $[HCO_3^-]$), and primary metabolic disturbances elicit predictable compensatory respiratory responses (secondary changes in $Paco_2$). Physiologic compensation can be predicted from the relationships displayed in **Table 55-1**. In general, with one exception, compensatory responses return the pH toward, but not to, the normal value. Chronic respiratory alkalosis when prolonged is an exception to this rule and may return the pH to a normal value. Metabolic acidosis due to an increase in endogenous acid production (e.g., ketoacidosis or lactic acid acidosis) lowers extracellular fluid $[HCO_3^-]$ and decreases extracellular pH. This stimulates the medullary chemoreceptors to increase ventilation and to return the ratio of $[HCO_3^-]$ to $Paco_2$, and thus pH, toward, but not typically to, the normal value. The degree of respiratory compensation expected in a metabolic acidosis can be predicted from the relationship: $Paco_2 = (1.5 \times [HCO_3^-]) + 8 \pm 2$ (Winter's equation). Thus, applying this equation, a patient with metabolic acidosis and $[HCO_3^-]$ of 12 mmol/L would be expected to have a $Paco_2$ of approximately 26 mmHg. In this example, if values for $Paco_2$ were <24 or >28 mmHg, values that exceed the boundaries for compensation for a simple disorder, a *mixed* disturbance should be recognized (metabolic acidosis plus respiratory alkalosis or metabolic acidosis plus respiratory acidosis, respectively). Compensatory responses for primary metabolic disorders move the $Paco_2$ in the same direction as the change in $[HCO_3^-]$, while compensation for primary respiratory disorders moves the $[HCO_3^-]$ in the same direction as the primary change in $Paco_2$ (Table 55-1). Therefore, changes in $Paco_2$ and $[HCO_3^-]$ in **opposite directions** (i.e., $Paco_2$ or $[HCO_3^-]$ is increased, whereas the other value is decreased) indicate a

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

acid-base values within the shaded area does not necessarily rule out a mixed disturbance. Imposition of one disorder over another may result in values lying within the area of a third. Thus, the nomogram, while convenient, is not a substitute for the equations in Table 55-1.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Therefore, changes in $Paco_2$ and $[HCO_3^-]$ in **opposite directions** (i.e., $Paco_2$ or $[HCO_3^-]$ is increased, whereas the other value is decreased) indicate a **mixed acid-base disturbance**. Another way to judge the appropriateness of the response in $[HCO_3^-]$ or $Paco_2$ is to use an acid-base nomogram (Fig. 55-1). While the shaded areas of the nomogram show the 95% confidence limits for physiologic compensation in simple disturbances, finding acid-base values within the shaded area does not necessarily rule out a mixed disturbance. Imposition of one disorder over another may result in values lying within the area of a third. Thus, the nomogram, while convenient, is not a substitute for the equations in Table 55-1.

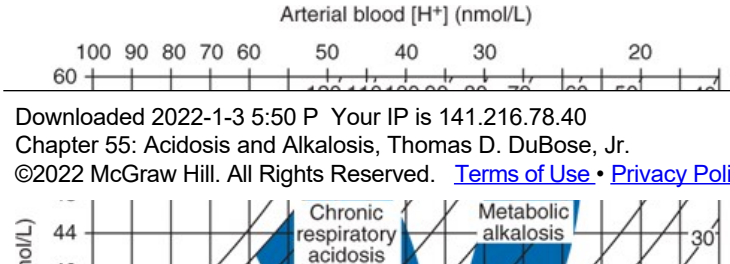
TABLE 55-1

Prediction of Compensatory Responses to Simple Acid-Base Disturbances and Pattern of Changes

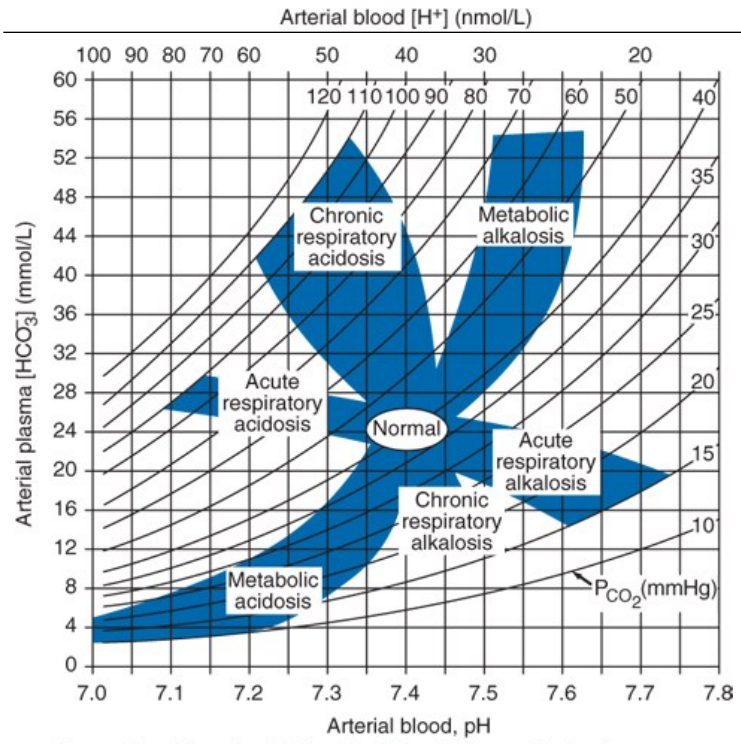
DISORDER	PREDICTION OF COMPENSATION	RANGE OF VALUES		
		pH	HCO_3^-	$Paco_2$
Metabolic acidosis	$Paco_2 = (1.5 \times HCO_3^-) + 8 \pm 2$ *or* $Paco_2$ will ↓ 1.25 mmHg per mmol/L ↓ in $[HCO_3^-]$ *or* $Paco_2 = [HCO_3^-] + 15$	Low	Low	Low
Metabolic alkalosis	$Paco_2$ will ↑ 0.75 mmHg per mmol/L ↑ in $[HCO_3^-]$ *or* $Paco_2$ will ↑ 6 mmHg per 10 mmol/L ↑ in $[HCO_3^-]$ *or* $Paco_2 = [HCO_3^-] + 15$	High	High	High
Respiratory alkalosis		High	Low	Low
Acute	$[HCO_3^-]$ will ↓ 0.2 mmol/L per mmHg ↓ in Pa_{CO_2}			
Chronic	$[HCO_3^-]$ will ↓ 0.4 mmol/L per mmHg ↓ in Pa_{CO_2}			
Respiratory acidosis		Low	High	High
Acute	$[HCO_3^-]$ will ↑ 0.1 mmol/L per mmHg ↑ in Pa_{CO_2}			
Chronic	$[HCO_3^-]$ will ↑ 0.4 mmol/L per mmHg ↑ in Pa_{CO_2}			

FIGURE 55-1

Acid-base nomogram. Shown are the 90% confidence limits (range of values) of the normal respiratory and metabolic compensations for primary acid-base disturbances. *(Reproduced with permission from LL Hamm and TD DuBose Jr, in Alan S.L. Yu, et al (eds): Brenner and Rector's The Kidney, 11th ed, pg. 505 Philadelphia, Elsevier, 2020.)*



Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

MIXED ACID-BASE DISORDERS

Mixed acid-base disorders—defined as independently coexisting disorders, not merely compensatory responses—are often seen in patients in critical care units and can lead to dangerous extremes of pH **(Table 55-2)**. The diagnosis of mixed acid-base disorders requires consideration of the anion gap (AG). To be accurate, the AG requires the presence of, or correction to, a normal serum albumin of 4.5 g/dL (see below, "Evaluate the Anion Gap"). If a patient with diabetic ketoacidosis (metabolic acidosis) and a high AG has an independent and concomitant respiratory disorder (e.g., pneumonia), the latter may lead to a superimposed respiratory acidosis or alkalosis and the $Paco_2$ will deviate from the predicted value for the response to a pure high-AG metabolic acidosis (Table 55-2). Patients with underlying chronic obstructive pulmonary disease may not respond to metabolic acidosis with an appropriate ventilatory response because of insufficient respiratory reserve (Table 55-2). Such imposition of respiratory acidosis on metabolic acidosis can lead to severe acidemia. When metabolic acidosis and metabolic alkalosis coexist in the same patient, the pH may be in the normal range. In this circumstance, it is the presence of an elevated AG (see below) that denotes the presence of a metabolic acidosis. Assuming a normal value for the AG of 10 mmol/L, incongruity in the ΔAG (existing minus normal AG) and the ΔHCO_3^- (normal value of 25 mmol/L minus abnormal HCO_3^- in the patient) indicates the presence of a mixed high-gap acidosis—metabolic alkalosis (see example below). A diabetic patient with ketoacidosis may have acute or chronic kidney failure resulting in a combination of metabolic acidoses from accumulation of both ketoacids and uremic acids. Patients who have ingested an overdose of drug combinations such as sedatives and salicylates may have mixed disturbances as a result of the acid-base response to the individual drugs (metabolic acidosis mixed with respiratory acidosis or respiratory alkalosis, respectively). Triple acid-base disturbances are more complex. For example, patients with metabolic acidosis due to alcoholic ketoacidosis may develop metabolic alkalosis due to vomiting and superimposed respiratory alkalosis due to the hyperventilation of hepatic dysfunction or alcohol withdrawal.

TABLE 55-2

Examples of Mixed Acid-Base Disorders

Mixed Metabolic and Respiratory
Metabolic acidosis—respiratory alkalosis
Key: High-AG metabolic acidosis; prevailing $Paco_2$ *below* predicted value (Table 55-1)
Metabolic acidosis—respiratory acidosis

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 55-2

Examples of Mixed Acid-Base Disorders

Mixed Metabolic and Respiratory
Metabolic acidosis—respiratory alkalosis
Key: High-AG metabolic acidosis; prevailing $PaCO_2$ *below* predicted value (Table 55-1)
Example: Na^+, 140; K^+, 4.0; Cl^-, 106; HCO_3^-, 14; AG, 20; Pa_{CO_2}, 24; pH, 7.39 (lactic acidosis, sepsis in ICU)
Metabolic acidosis—respiratory acidosis
Key: High-AG metabolic acidosis; prevailing $PaCO_2$ *above* predicted value (Table 55-1)
Example: Na^+, 140; K^+, 4.0; Cl^-, 102; HCO_3^-, 18; AG, 20; Pa_{CO_2}, 38; pH, 7.30 (severe pneumonia, pulmonary edema)
Metabolic alkalosis—respiratory alkalosis
Key: $PaCO_2$ does not increase as predicted; pH higher than expected
Example: Na^+, 140; K^+, 4.0; Cl^-, 91; HCO_3^-, 33; AG, 16; Pa_{CO_2}, 38; pH, 7.55 (liver disease and diuretics)
Metabolic alkalosis—respiratory acidosis
Key: $PaCO_2$ higher than predicted; pH normal
Example: Na^+, 140; K^+, 3.5; Cl^-, 88; HCO_3^-, 42; AG, 10; Pa_{CO_2}, 67; pH, 7.42 (COPD on diuretics)
Mixed Metabolic Disorders
Metabolic acidosis—metabolic alkalosis
Key: Only detectable with high-AG acidosis; ΔAG (10) >> $\Delta HCO_3^{-(o)}$
Example: Na^+, 140; K^+, 3.0; Cl^-, 95; HCO_3^-, 25; AG, 20; Pa_{CO_2}, 40; pH, 7.42 (uremia with vomiting)
Metabolic acidosis—metabolic acidosis
Key: Mixed high-AG—normal-AG acidosis; ΔHCO_3^- accounted for by combined change in ΔAG and ΔCl^-
Example: Na^+, 135; K^+, 3.0; Cl^-, 110; HCO_3^-, 10; AG, 15; Pa_{CO_2}, 25; pH, 7.20 (diarrhea and lactic acidosis, toluene toxicity, treatment of diabetic ketoacidosis)

Abbreviations: AG, anion gap; COPD, chronic obstructive pulmonary disease; ICU, intensive care unit.

APPROACH TO THE PATIENT

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

simultaneously, prior to therapy. An increase in $[HCO_3^-]$ occurs with either metabolic alkalosis or respiratory acidosis. Conversely, a decrease in $[HCO_3^-]$ occurs with either metabolic acidosis or respiratory alkalosis. In the determination of arterial blood gases by the clinical laboratory, both pH

APPROACH TO THE PATIENT

Acid-Base Disorders

The diagnosis of acid-base disorders follows a stepwise approach **(Table 55-3)**. Blood for electrolytes and arterial blood gases should be drawn simultaneously, prior to therapy. An increase in [HCO_3^-] occurs with either metabolic alkalosis or respiratory acidosis. Conversely, a decrease in [HCO_3^-] occurs with either metabolic acidosis or respiratory alkalosis. In the determination of arterial blood gases by the clinical laboratory, both pH and $Paco_2$ are measured, and the [HCO_3^-] is calculated from the Henderson-Hasselbalch equation. This *calculated* value should be compared with the *measured* [HCO_3^-] (or total CO_2) on the electrolyte panel. These two values should agree within 2 mmol/L. If they do not, the values may not have been drawn simultaneously, or a laboratory error may be present. After verifying the blood acid-base values, the precise acid-base disorder can then be classified.

EVALUATE THE ANION GAP

Evaluations of acid-base disorders should involve acknowledgement of the AG. The AG is calculated, either by the clinical laboratory or the clinician, as follows: AG = $Na^+ - (Cl^- + HCO_3^-)$. The value for plasma [K^+] is typically omitted from the calculation of the AG in the United States. The "normal" value for the AG reported by clinical laboratories has declined with improved methodology for measuring plasma electrolytes and ranges from 6–12 mmol/L, with an average of approximately 10 mmol/L. The unmeasured anions normally present in plasma include anionic proteins (e.g., albumin), phosphate, sulfate, and organic anions. When acid anions, such as acetoacetate and lactate, accumulate in extracellular fluid, the AG increases, causing a **high-AG acidosis**. An increase in the AG is most often due to an increase in unmeasured anions but, less commonly, may be due to a decrease in unmeasured cations (calcium, magnesium, potassium). In addition, the AG may increase with an increase in anionic albumin (e.g., severe dehydration). A decrease in the AG can be due to (1) an increase in unmeasured cations; (2) the addition to the blood of abnormal cations, such as lithium (lithium intoxication) or cationic immunoglobulins (plasma cell dyscrasias); (3) a reduction in the plasma anion albumin concentration (nephrotic syndrome, liver disease, or malabsorption); or (4) hyperviscosity and severe hyperlipidemia, which can lead to an underestimation of sodium and chloride concentrations. Because the normal AG of 10 mmol/L assumes that the serum albumin is normal, if hypoalbuminemia is present, the value for the AG must be corrected. For example, for each g/dL of serum albumin below the normal value (4.5 g/dL), 2.5 mmol/L should be added to the reported (uncorrected) AG. Thus, in a patient with a serum albumin of 2.5 g/dL (2 g/dL below the normal value) and an uncorrected AG of 15, the corrected AG is calculated by adding 5 mmol/L (2.5 × 2 = 5; 5 + 15 = corrected AG of 20 mmol/L). Clinical laboratories do not correct the AG for coexisting hypoalbuminemia and typically report the uncorrected value, requiring the attention of the clinician to the prevailing serum albumin concentration. The clinical disorders that may cause a high-AG acidosis are displayed in Table 55-3.

A high AG is usually due to accumulation of non–chloride-containing acids that contain inorganic (phosphate, sulfate), organic (ketoacids, lactate, uremic organic anions), exogenous (salicylate or ingested toxins with organic acid production), or unidentified anions. The high AG is meaningful even if the [HCO_3^-] or pH is normal. Simultaneous metabolic acidosis of the high-AG variety plus either chronic respiratory acidosis or metabolic alkalosis represents a situation in which [HCO_3^-] may be normal or even high (Table 55-3). In cases of high-AG metabolic acidosis, it is valuable to compare the decline in [HCO_3^-] (ΔHCO_3^-: 25 – patient's [HCO_3^-]) with the increase in the AG (ΔAG: patient's AG – 10).

Similarly, normal values for [HCO_3^-], $Paco_2$, and pH do not ensure the absence of an acid-base disturbance. For instance, an alcoholic who has been vomiting may develop a metabolic alkalosis with a pH of 7.55, $Paco_2$ of 47 mmHg, [HCO_3^-] of 40 mmol/L, [Na^+] of 135, [Cl^-] of 80, and [K^+] of 2.8. If such a patient were then to develop a superimposed alcoholic ketoacidosis with a β-hydroxybutyrate concentration of 15 mmol/L, arterial pH would fall to 7.40, the [HCO_3^-] to 25 mmol/L, and the $Paco_2$ to 40 mmHg. Although these blood gases are normal, the AG is elevated at 30 mmol/L, indicating a mixed metabolic alkalosis and metabolic acidosis is present. A mixture of high-gap acidosis and metabolic alkalosis is recognized easily by comparing the differences (Δ values) in the normal to prevailing patient values. In this example, the ΔHCO_3^- is 0 (25 − 25 mmol/L), but the ΔAG is 20 (30 – 10 mmol/L). Therefore, 20 mmol/L is unaccounted for in the Δ/Δ value (ΔAG to ΔHCO_3^-).

TABLE 55-3

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

1. Obtain arterial blood gas (ABG) and electrolytes simultaneously.
2. Compare [HCO_3^-] on ABG and electrolytes to verify accuracy.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 55-3

Steps in Acid-Base Diagnosis

1. Obtain arterial blood gas (ABG) and electrolytes simultaneously.
2. Compare $[HCO_3^-]$ on ABG and electrolytes to verify accuracy.
3. Evaluate anion gap (AG); if not normal, correct to albumin concentration of 4.5 g/dL (see text).
4. Know four causes of high-AG acidosis (ketoacidosis, lactic acid acidosis, renal failure, and toxins).
5. Know two causes of hyperchloremic or nongap acidosis (bicarbonate loss from gastrointestinal tract, renal tubular acidosis).
6. Estimate compensatory response (Table 55-1).
7. Compare ΔAG and ΔHCO_3^-.
8. Compare change in $[Cl^-]$ with change in $[Na^+]$.

METABOLIC ACIDOSIS

Metabolic acidosis can occur because of an increase in endogenous acid production (such as lactate and ketoacids), loss of bicarbonate (as in diarrhea), or accumulation of endogenous acids because of inappropriately low excretion of net acid by the kidney (as in chronic kidney disease). Metabolic acidosis has profound effects on the respiratory, cardiac, and nervous systems. The fall in blood pH is accompanied by a characteristic increase in ventilation, especially the tidal volume (Kussmaul respiration). Intrinsic cardiac contractility may be depressed, but inotropic function can be normal because of catecholamine release. Both peripheral arterial vasodilation and central venoconstriction may be present; the decrease in central and pulmonary vascular compliance predisposes to pulmonary edema with even minimal volume overload. CNS function is depressed, with headache, lethargy, stupor, and, in some cases, even coma. Glucose intolerance may also occur.

There are two major categories of clinical metabolic acidosis: high-AG and non-AG acidosis (Table 55-3 and **Table 55-4**). The presence of metabolic acidosis, a normal AG, and hyperchloremia denotes the presence of a non-AG metabolic acidosis.

TABLE 55-4

Causes of High-Anion Gap Metabolic Acidosis

Lactic acidosis	Toxins
Ketoacidosis	Ethylene glycol
Diabetic	Methanol
Alcoholic	Salicylates
Starvation	Propylene glycol
	Pyroglutamic acid (5-oxoproline)
	Renal failure (acute and chronic)

TREATMENT

Metabolic Acidosis

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

is metabolizable (i.e., β-hydroxybutyrate, acetoacetate, and lactate). Conversely, nonmetabolizable anions that may accumulate in advanced-stage

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Metabolic Acidosis

Treatment of metabolic acidosis with alkali should be reserved for severe acidemia except when the patient has no "potential HCO_3^-" in plasma. The potential [HCO_3^-] can be estimated from the increment (Δ) in the AG (ΔAG = patient's AG – 10), only if the acid anion that has accumulated in plasma is metabolizable (i.e., β-hydroxybutyrate, acetoacetate, and lactate). Conversely, nonmetabolizable anions that may accumulate in advanced-stage chronic kidney disease or after toxin ingestion are not metabolizable and do not represent "potential" HCO_3^-. In patients with acute kidney failure or acute-on-chronic kidney failure, improvement in kidney function after volume resuscitation may improve the serum [HCO_3^-], but this is a slow and unpredictable process. Consequently, patients with a non-AG acidosis (hyperchloremic acidosis) or an AG acidosis attributable to a nonmetabolizable anion due to advanced kidney failure ("uremic" acidosis) should receive alkali therapy, either PO ($NaHCO_3$ tablets or Shohl's solution) or IV ($NaHCO_3$), in an amount necessary to slowly increase the plasma [HCO_3^-] to a target value of 22 mmol/L. Importantly, overcorrection should be avoided.

Bicarbonate therapy in diabetic ketoacidosis (DKA) is reserved for adult patients with severe acidemia (pH <7.00) and/or evidence of shock. In such circumstances, bicarbonate may be administered IV, as a slow infusion of 50 meq of $NaHCO_3$ diluted in 300 mL of a saline solution, over 30–45 min, during the initial 1–2 h of therapy. Bolus administration should be avoided. Administration of $NaHCO_3$ requires careful monitoring of plasma electrolytes during the course of therapy because of the risk for hypokalemia as urine output is established. A reasonable initial goal in DKA is to increase the [HCO_3^-] to 10–12 mmol/L and the pH to approximately 7.20, but clearly not to increase these values to normal.

HIGH-ANION GAP ACIDOSES

APPROACH TO THE PATIENT

High-Anion Gap Acidoses

There are four principal causes of a high-AG acidosis: (1) lactic acidosis, (2) ketoacidosis, (3) ingested toxins, and (4) acute and chronic kidney failure (Table 55-4). Initial screening to differentiate the high-AG acidoses should include (1) a probe of the history for evidence of drug and toxin ingestion and measurement of arterial blood gas to detect coexistent respiratory alkalosis (salicylates); (2) determination of whether a history of diabetes mellitus is present (DKA); (3) a search for evidence of alcoholism or increased levels of β-hydroxybutyrate (alcoholic ketoacidosis); (4) observation for clinical signs of uremia and determination of the blood urea nitrogen (BUN) and creatinine (uremic acidosis); (5) inspection of the urine for oxalate crystals (ethylene glycol ingestion); and (6) recognition of the numerous clinical settings in which lactate levels may be increased (hypotension, shock, cardiac failure, leukemia, cancer, and drug or toxin ingestion).

Lactic Acidosis

An increase in plasma L-lactate may be secondary to poor tissue perfusion (type A)—circulatory insufficiency (shock, cardiac failure), severe anemia, mitochondrial enzyme defects, and inhibitors (carbon monoxide, cyanide)—or to aerobic disorders (type B)—malignancies, nucleoside analogue reverse transcriptase inhibitors in HIV, diabetes mellitus, renal or hepatic failure, thiamine deficiency, severe infections (cholera, malaria), seizures, or drugs/toxins (biguanides, ethanol, and the toxic alcohols: ethylene glycol, methanol, or propylene glycol). Unrecognized bowel ischemia or infarction in a patient with severe atherosclerosis or cardiac decompensation receiving vasopressors is a common cause of lactic acidosis in elderly patients. Pyroglutamic acidemia may occur in critically ill patients receiving acetaminophen, which causes depletion of glutathione and accumulation of 5-oxyprolene. D-Lactic acid acidosis, which may be associated with jejunoileal bypass, short bowel syndrome, or intestinal obstruction, is due to formation of D-lactate by gut bacteria.

APPROACH TO THE PATIENT

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The overarching goal of treatment is to correct the underlying condition that disrupts lactate metabolism; tissue perfusion should be restored when inadequate, but vasoconstrictors should be avoided, or used cautiously, because they may worsen tissue perfusion. Alkali therapy is generally

APPROACH TO THE PATIENT

L-Lactic Acid Acidosis

The overarching goal of treatment is to correct the underlying condition that disrupts lactate metabolism; tissue perfusion should be restored when inadequate, but vasoconstrictors should be avoided, or used cautiously, because they may worsen tissue perfusion. Alkali therapy is generally advocated for acute, severe acidemia (pH <7.00) to improve cardiovascular function. However, $NaHCO_3$ therapy may paradoxically depress cardiac performance and exacerbate acidosis by enhancing lactate production (HCO_3^- stimulates phosphofructokinase). While the use of alkali in moderate lactic acidosis is controversial, it is generally agreed that attempts to return the pH or [HCO_3^-] to normal by administration of exogenous $NaHCO_3$ are deleterious. A reasonable approach with severe acidemia is to infuse sufficient $NaHCO_3$ to raise arterial pH to no more than 7.2 or the [HCO_3^-] to no more than 12 mmol/L.

$NaHCO_3$ therapy can cause fluid overload, hypercapnia, and hypertension because the amount required can be massive when accumulation of lactic acid is relentless. Fluid administration is poorly tolerated, especially in the oliguric patient, when central venoconstriction coexists. If the underlying cause of the lactic acidosis can be remedied, blood lactate will be converted to HCO_3^- and may result in an overshoot alkalosis if exogenous $NaHCO_3$ has been administered excessively.

Ketoacidosis

DIABETIC KETOACIDOSIS (DKA)

This condition is caused by increased fatty acid metabolism and the accumulation of ketoacids (acetoacetate and β-hydroxybutyrate). DKA usually occurs in insulin-dependent diabetes mellitus in association with cessation of insulin or an intercurrent illness such as an infection, gastroenteritis, pancreatitis, or myocardial infarction, which increases insulin requirements temporarily and acutely, and is characterized by hyperglycemia, ketonemia, and a high-AG acidosis. Nevertheless, the plasma glucose may be normal or only slightly elevated in the setting of starvation ketoacidosis or in diabetics receiving antagonists of the proximal tubule sodium-glucose co-transporter 2 (SGLT2). These agents cause glycosuria, an osmotic diuresis, and lower the plasma glucose. Ketoacidosis can occur in patients receiving SGLT2 antagonists for the same reasons as in classical DKA, but the plasma glucose is typically normal or only slightly elevated. The accumulation of ketoacids in plasma accounts for the increment in the AG in both classical DKA and euglycemic DKA. Measurement of urine ketones (by the dipstick nitroprusside reaction) does not detect β-hydroxybutyrate and may underestimate the degree of ketosis (see below). Excretion of ketoacids obligates the excretion of cations, such as Na^+ and K^+, contributing to volume depletion and Cl^- retention. In some circumstances, a mixed non-AG–high-AG acidosis may occur simultaneously and is recognized when the ΔHCO_3^- exceeds the ΔAG. It should be noted that bicarbonate therapy is rarely necessary in DKA except with extreme acidemia (pH <7.00) or if the patient is in shock. If administered, $NaHCO_3$ should be administered in only limited amounts because of the risk for cerebral edema. Patients with DKA are typically volume depleted and require fluid resuscitation with isotonic saline. Volume overexpansion should be avoided, however, because overly aggressive saline administration may cause hyperchloremic acidosis during or following treatment of DKA. Regular insulin should be administered IV as an initial bolus of 0.1 U/kg followed by an infusion of 0.1 U/kg/h until the AG returns to normal; see **Chap. 403** for more detail.

ALCOHOLIC KETOACIDOSIS (AKA)

AKA is usually associated with chronic alcoholism, binge drinking, vomiting, abdominal pain, poor nutrition, and volume depletion. The glucose concentration is variable, and acidosis may be severe because of elevated ketones, predominantly β-hydroxybutyrate. The presence of a high-AG acidosis, in the absence of hyperglycemia, in a patient with chronic alcoholism suggests the diagnosis of AKA. Mixed acid-base disorders are common in AKA. Hypoperfusion may enhance lactic acid production (mixed high-AG acidosis), chronic respiratory alkalosis may accompany liver disease (mixed high-AG acidosis and respiratory alkalosis), and metabolic alkalosis can result from vomiting (mixed high-AG acidosis and metabolic alkalosis: ΔAG exceeds ΔHCO_3^-). As the circulation is restored by administration of IV fluids, the preferential accumulation of β-hydroxybutyrate is then shifted to acetoacetate. This explains the common clinical observation of an increasingly positive nitroprusside reaction (ketones) as the circulation is restored. The nitroprusside reaction can detect acetoacetic acid but not β-hydroxybutyrate, so that the degree of ketosis and ketonuria can not only change with may have a mixed high-AG–non-AG metabolic acidosis (ΔHCO_3^- exceeds ΔAG).

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

acetoacetate. This explains the common clinical observation of an increasingly positive nitroprusside reaction (ketones) as the circulation is restored. The nitroprusside reaction can detect acetoacetic acid but not β-hydroxybutyrate, so that the degree of ketosis and ketonuria can not only change with therapy, but can be underestimated initially. Therefore, the plasma β-hydroxybutyrate level should be measured. Patients with AKA usually present with relatively normal renal function, as opposed to DKA, where renal function is often compromised because of volume depletion (osmotic diuresis) or diabetic nephropathy. The AKA patient with normal renal function may excrete relatively large quantities of ketoacids and retain Cl^- and, therefore, may have a mixed high-AG–non-AG metabolic acidosis (ΔHCO_3^- exceeds ΔAG).

TREATMENT

Alcoholic Ketoacidosis

Extracellular fluid deficits almost always accompany AKA and should be repaired by IV administration, initially, of saline and glucose (5% dextrose in 0.9% NaCl). Hypophosphatemia, hypokalemia, and hypomagnesemia may coexist and should be monitored carefully and corrected when indicated. Hypophosphatemia may emerge 12–24 h after admission, exacerbated by glucose infusion, and, if severe, may induce marked muscle weakness, hemolysis, rhabdomyolysis, or respiratory arrest. Upper gastrointestinal hemorrhage, pancreatitis, and pneumonia may accompany this disorder.

Drug- and Toxin-Induced Acidosis

SALICYLATES

(See also Chap. 458) Salicylate intoxication in adults usually causes respiratory alkalosis or a mixture of high-AG metabolic acidosis and respiratory alkalosis. Only a portion of the AG is due to salicylates. Lactic acid production is also often increased.

TREATMENT

Salicylate-Induced Acidosis

Vigorous gastric lavage with isotonic saline (not $NaHCO_3$) should be initiated immediately. All patients should receive at least one round of activated charcoal per nasogastric tube (1 g/kg up to 50 g). To facilitate excretion of salicylate in the acidotic patient, IV $NaHCO_3$ is administered in amounts adequate to alkalinize the urine (urine pH >7.5) and to maintain urine output. Raising urine pH from 6.5 to 7.5 increases salicylate clearance fivefold. Patients with coexisting respiratory alkalosis should also receive $NaHCO_3$ cautiously to avoid excessive alkalemia. Acetazolamide may be administered in the face of alkalemia, when an alkaline diuresis cannot be achieved, or to ameliorate volume overload associated with $NaHCO_3$ administration. Acetazolamide may cause systemic metabolic acidosis if the excreted HCO_3^- is not replaced, a circumstance that can markedly reduce salicylate clearance. **Hypokalemia should be anticipated** with vigorous bicarbonate therapy and should be treated promptly and aggressively. Glucose-containing fluids should be administered because of the danger of hypoglycemia. Excessive insensible fluid losses may cause severe volume depletion and hypernatremia. If renal failure prevents rapid clearance of salicylate, hemodialysis should be performed against a standard bicarbonate dialysate ($[HCO_3^-]$ = 30–35 meq/L).

ALCOHOLS

Under most physiologic conditions, sodium, urea, and glucose generate the osmotic pressure of blood. Plasma osmolality is calculated according to the following expression: $P_{osm} = 2Na^+ + Glu + BUN$ (all in mmol/L), or, using conventional laboratory values in which glucose and BUN are expressed in mg/dL: $P_{osm} = 2Na^+ + Glu/18 + BUN/2.8$. The calculated and determined osmolality should agree within 10–15 mmol/kg H_2O. When the measured osmolality exceeds the calculated osmolality by >10–15 mmol/kg H_2O, one of two circumstances prevails. Either the serum sodium is spuriously low, as with hyperlipidemia or hyperproteinemia (pseudohyponatremia), or osmolytes other than sodium salts, glucose, or urea have accumulated in plasma. Examples of such osmolytes include mannitol, radiocontrast media, ethanol, isopropyl alcohol, ethylene glycol, propylene glycol, methanol, and acetone. In this situation, the difference between the calculated osmolality and the measured osmolality (*osmolar gap*) is proportional to the

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

unless extreme overdose causes hypotension and lactic acid acidosis.

acetone. In this situation, the difference between the calculated osmolality and the measured osmolality (*osmolar gap*) is proportional to the concentration of the unmeasured solute. With an appropriate clinical history and index of suspicion, identification of an osmolar gap is helpful in identifying the presence of toxic alcohol-associated AG acidosis. Three alcohols may cause fatal intoxications: ethylene glycol, methanol, and isopropyl alcohol. All cause an elevated osmolal gap, but only the first two cause a high-AG acidosis. Isopropyl alcohol ingestion does not typically elevate the AG unless extreme overdose causes hypotension and lactic acid acidosis.

ETHYLENE GLYCOL

(See also Chap. 458) Ethylene glycol (EG) (commonly used in antifreeze, but also in brake fluid and windshield washer fluid deicers) is metabolized by alcohol dehydrogenase, and ingestion of EG leads to a metabolic acidosis and severe damage to the CNS, heart, lungs, and kidneys. The combination of both a high AG and osmolar gap is highly suspicious for EG or methanol intoxication. The combination of a high AG and high osmolar gap in a patient suspected of EG ingestion should be taken as evidence of EG toxicity prior to measurement of EG levels, and treatment should not be delayed. The osmolar gap may be elevated earlier than the AG, and as the osmolar gap declines, the AG increases. The increased AG and osmolar gap in EG intoxication are attributable to EG and its metabolites, glycolate, oxalate, and other organic acids. Lactic acid production increases secondary to inhibition of the tricarboxylic acid cycle and altered intracellular redox state and may contribute to the high AG. Acute tubule injury is caused initially by glycolate and later is amplified by tubule obstruction from oxalate crystals.

TREATMENT

Ethylene Glycol Intoxication

This includes the prompt institution of IV isotonic fluids, thiamine and pyridoxine supplements, fomepizole, and usually, hemodialysis. Both fomepizole and ethanol compete with EG for metabolism by alcohol dehydrogenase. Fomepizole (4-methylpyrazole; 15 mg/kg IV over 30 min as a loading dose, then 10 mg/kg for four doses every 12 h) is the agent of choice and offers the advantages of a predictable decline in EG levels without excessive obtundation, as seen during ethyl alcohol infusion. Fomepizole should be continued until blood pH is normal or the osmolar gap is <10 mOsm/kg H_2O. Hemodialysis is indicated when the arterial pH is <7.3, a high-AG acidosis is present, the osmolar gap exceeds 20 mOsm/kg H_2O, or there is evidence of end organ damage such as CNS manifestations and kidney failure.

METHANOL

(See also Chap. 458) The ingestion of methanol (wood alcohol) causes metabolic acidosis, and its metabolites formaldehyde and formic acid cause severe optic nerve and CNS damage. Lactic acid, ketoacids, and other unidentified organic acids may contribute to the acidosis. Due to its low molecular mass (32 Da), an osmolar gap is present and may precede the elevation of the AG.

TREATMENT

Methanol Intoxication

Treatment of methanol intoxication is similar to that for EG intoxication, including general supportive measures, fomepizole, and hemodialysis.

PROPYLENE GLYCOL

Propylene glycol is the vehicle used in IV administration of diazepam, lorazepam, phenobarbital, nitroglycerine, etomidate, enoximone, and phenytoin. Propylene glycol is generally safe for limited use in these IV preparations, but toxicity has been reported in the setting of the intensive care unit in patients receiving frequent or continuous therapy, where the propylene glycol vehicle may accumulate in the plasma. This form of high-gap acidosis should be considered in patients with unexplained high-gap acidosis, hyperosmolality, and clinical deterioration, especially in the setting of treatment for alcohol withdrawal. Propylene glycol, like EG and methanol, is metabolized by alcohol dehydrogenase. With intoxication by propylene glycol, the first response is to stop the offending infusion. Additionally, fomepizole should also be administered in acidotic patients.

ISOPROPYL ALCOHOL

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

acetone is rapidly excreted. Both isopropyl alcohol and acetone increase the osmolar gap, and hypoglycemia is common. Alternative diagnoses should be considered if the patient does not improve significantly within a few hours. Patients with hemodynamic instability with plasma levels above 400

Ingested isopropanol is absorbed rapidly and may be fatal when as little as 150 mL of rubbing alcohol, solvent, or deicer is consumed. A plasma level >400 mg/dL is life-threatening. Isopropyl alcohol is metabolized by alcohol dehydrogenase to acetone. The characteristic features differ significantly from EG and methanol intoxication in that the parent compound, not the metabolites, causes toxicity, and a high-AG acidosis is ***not*** present because acetone is rapidly excreted. Both isopropyl alcohol and acetone increase the osmolar gap, and hypoglycemia is common. Alternative diagnoses should be considered if the patient does not improve significantly within a few hours. Patients with hemodynamic instability with plasma levels above 400 mg/dL should be considered for hemodialysis.

TREATMENT

Isopropyl Alcohol Toxicity

Isopropanol alcohol toxicity is treated by supportive therapy, IV fluids, pressors, ventilatory support if needed, and acute hemodialysis for prolonged coma, hemodynamic instability, or levels >400 mg/dL.

PYROGLUTAMIC ACID

Acetaminophen-induced high-AG metabolic acidosis is uncommon but is recognized in either patients with acetaminophen overdose or malnourished or critically ill patients receiving acetaminophen in typical dosage. 5-Oxoproline accumulation after acetaminophen should be suspected in the setting of an unexplained high-AG acidosis without elevation of the osmolar gap in patients receiving acetaminophen. The first step in treatment is to immediately discontinue acetaminophen. Additionally, sodium bicarbonate IV should be given. Although *N*-acetylcysteine has been suggested, it is not proven that it hastens the metabolism of 5-oxoproline by increasing intracellular glutathione concentrations in this setting, as assumed.

Chronic Kidney Disease

(See also Chap. 311) The hyperchloremic acidosis of moderate chronic kidney disease (CKD; stage 3) is eventually converted to the high-AG acidosis of advanced renal failure (stages 4 and 5 CKD). Poor filtration and reabsorption of organic anions contribute to the pathogenesis. As renal disease progresses, the number of functioning nephrons eventually becomes insufficient to keep pace with net acid production. Uremic acidosis in advanced CKD is characterized, therefore, by a reduced rate of NH_4^+ production and excretion. Alkaline salts from bone buffer the acid retained in CKD. Despite significant retention of acid (up to 20 mmol/d), the serum [HCO_3^-] does not typically decrease further, indicating participation of buffers outside the extracellular compartment. Therefore, the trade-off in untreated chronic metabolic acidosis of CKD stages 3 and 4 is significant loss of bone mass due to reduction in bone calcium carbonate. Chronic acidosis also contributes significantly to muscle wasting and disability in advancing CKD.

TREATMENT

Metabolic Acidosis of Chronic Kidney Disease

Because of the association of metabolic acidosis in advanced CKD with muscle catabolism, bone disease, and more rapid progression of CKD, both the “uremic acidosis” of end-stage renal disease and the non-AG metabolic acidosis of stages 3 and 4 CKD require oral alkali replacement to increase and maintain the [HCO_3^-] to a value >22 mmol/L. This can be accomplished with relatively modest amounts of alkali (1.0–1.5 mmol/kg body weight per day) and has been shown to slow the progression of CKD. Either $NaHCO_3$ tablets (650-mg tablets contain 7.8 meq) or oral sodium citrate (Shohl’s solution) is effective. Moreover, addition of fruits and vegetables (citrate) to the diet may increase the plasma [HCO_3^-] and slow progression.

NON–ANION GAP METABOLIC ACIDOSES

Alkali can be lost from the gastrointestinal tract as a result of diarrhea or from the kidneys due to renal tubular abnormalities (e.g., renal tubular acidosis [RTA]). In these disorders **(Table 55-5)**, reciprocal changes in [Cl^-] and [HCO_3^-] result in a normal AG. In non–AG acidosis, therefore, the

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Causes of Non–Anion Gap Acidosis

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

acidosis [RTA]). In these disorders **(Table 55-5)**, reciprocal changes in [Cl^-] and [HCO_3^-] result in a normal AG. In non–AG acidosis, therefore, the increase in [Cl^-] above the normal value approximates the decrease in [HCO_3^-]. The absence of such a relationship suggests a mixed disturbance.

TABLE 55-5

Causes of Non–Anion Gap Acidosis

I. Gastrointestinal bicarbonate loss
  A. Diarrhea
  B. External pancreatic or small-bowel drainage
  C. Ureterosigmoidostomy, jejunal loop, ileal loop
  D. Drugs
    1. Calcium chloride (acidifying agent)
    2. Magnesium sulfate (diarrhea)
    3. Cholestyramine (bile acid diarrhea)

II. Renal acidosis
  A. Hypokalemia
    1. Proximal RTA (type 2)
      Drug-induced: acetazolamide, topiramate
    2. Distal (classic) RTA (type 1)
      Drug-induced: amphotericin B, ifosfamide
  B. Hyperkalemia
    1. Generalized distal nephron dysfunction (type 4 RTA)
      a. Selective aldosterone deficiency
      b. Mineralocorticoid resistance (PHA I, autosomal dominant)
      c. Voltage defect (PHA I, autosomal recessive, and PHA II)
      d. Hyporeninemic hypoaldosteronism
      e. Tubulointerstitial disease
  C. Normokalemia
    1. Chronic progressive kidney disease

III. Drug-induced hyperkalemia (with renal insufficiency)
  A. Potassium-sparing diuretics (amiloride, triamterene, spironolactone, eplerenone)
  B. Trimethoprim
  C. Pentamidine
  D. ACE-Is and ARBs
  E. Nonsteroidal anti-inflammatory drugs
  F. Calcineurin inhibitors
  G. Heparin in critically ill patients

IV. Other
  A. Acid loads (ammonium chloride, hyperalimentation)
  B. Loss of potential bicarbonate: ketosis with ketone excretion
  C. Expansion acidosis (rapid saline administration)
  D. Hippurate
  E. Cation exchange resins

Abbreviations: ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; PHA, pseudohypoaldosteronism; RTA, renal tubular acidosis.

Stool contains a higher concentration of HCO_3^- and decomposed HCO_3^- than plasma so that metabolic acidosis develops in diarrhea. Instead of an acid urine pH (as anticipated with systemic acidosis), urine pH is usually >6 because metabolic acidosis and hypokalemia increase renal synthesis and

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

by clinical laboratories but can be estimated by calculating the urine anion gap (UAG): UAG = $[Na^+ + K^+]_u - [Cl^-]_u$. When $[Cl^-]_u > [Na^+ + K^+]_u$, the UAG is negative by definition. This suggests that the urine ammonium level is appropriately increased, suggesting an extrarenal cause of the acidosis.

acid urine pH (as anticipated with systemic acidosis), urine pH is usually >6 because metabolic acidosis and hypokalemia increase renal synthesis and excretion of NH_4^+, thus providing a urinary buffer that increases urine pH. Metabolic acidosis due to gastrointestinal losses with a high urine pH can be differentiated from RTA because urinary NH_4^+ excretion is typically low in RTA and high with diarrhea. Urinary NH_4^+ levels are not routinely measured by clinical laboratories but can be estimated by calculating the urine anion gap (UAG): UAG = $[Na^+ + K^+]_u - [Cl^-]_u$. When $[Cl^-]_u > [Na^+ + K^+]_u$, the UAG is negative by definition. This suggests that the urine ammonium level is appropriately increased, suggesting an extrarenal cause of the acidosis. Conversely, when the UAG is positive, the urine ammonium level is predictably low, suggesting a renal tubular origin of the acidosis. Recent studies have shown a poor correlation between the UAG and the measured urine ammonium, thus calling the estimation of urine ammonium by calculation of the UAG into question. Therefore, clinical laboratories should be encouraged to measure urine ammonium by adaptation of automated plasma ammonium assays, using the enzymatic method, if the urine sample is diluted 1:200 in normal saline.

Proximal RTA (type 2 RTA) **(Chap. 315)** is most often due to generalized proximal tubular dysfunction manifested by glycosuria, generalized aminoaciduria, and phosphaturia (Fanconi syndrome). When the plasma $[HCO_3^-]$ is low, the urine pH is acid (pH <5.5) but exceeds 5.5 with alkali therapy. The fractional excretion of $[HCO_3^-]$ may exceed 10–15% when the serum HCO_3^- is >20 mmol/L. Because of the defect in HCO_3^- reabsorption by the proximal tubule, therapy with $NaHCO_3$ will enhance delivery of HCO_3^- to the distal nephron and enhance renal potassium secretion, thereby causing hypokalemia.

The typical findings in acquired or inherited forms of **classic distal RTA** (type 1 RTA) include hypokalemia, a non-AG metabolic acidosis, low urinary NH_4^+ excretion (positive UAG, low urine $[NH_4^+]$), and inappropriately high urine pH (pH >5.5). Most patients have hypocitraturia and hypercalciuria; nephrolithiasis, nephrocalcinosis, and bone disease are common. In **generalized distal RTA** (type 4 RTA), hyperkalemia is disproportionate to the reduction in glomerular filtration rate (GFR) because of coexisting dysfunction of potassium and acid secretion. Urinary ammonium excretion is invariably depressed, and kidney function may be compromised secondary to diabetic nephropathy, obstructive uropathy, or chronic tubulointerstitial disease.

Hyporeninemic hypoaldosteronism typically presents as a non-AG metabolic acidosis in older adults with diabetes mellitus or tubulointerstitial disease and CKD (estimated GFR 20–50 mL/min) with hyperkalemia ($[K^+]$ 5.2–6.0 mmol/L), concurrent hypertension, and congestive heart failure. Both the metabolic acidosis and the hyperkalemia are out of proportion to impairment in GFR. Nonsteroidal anti-inflammatory drugs, trimethoprim, pentamidine, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin receptor blockers (ARBs) can also increase the risk for hyperkalemia and a non-AG metabolic acidosis in patients with CKD, especially from diabetic nephropathy (Table 55-5).

TREATMENT

Non–Anion Gap Metabolic Acidoses

For non-AG acidosis due to gastrointestinal losses of bicarbonate, $NaHCO_3$ may be administered intravenously or orally, as determined by the severity of both the acidosis and the accompanying volume depletion. Proximal RTA is the most challenging of the RTAs to treat if the goal is to restore the serum $[HCO_3^-]$ to normal because administration of oral alkali increases urinary excretion of bicarbonate and potassium. In patients with proximal RTA (type 2), potassium administration is typically required. An oral solution of sodium and potassium citrate (citric acid 334 mg, sodium citrate 500 mg, and potassium citrate 550 mg per 5 mL) may be prescribed for this purpose (Virtrate or Cytra-3). In classical distal RTA (type 1), hypokalemia should be corrected first. When accomplished, alkali therapy with either sodium citrate (Shohl's solution) or $NaHCO_3$ tablets (650-mg tablets contain 7.8 meq) should be initiated to correct and maintain the serum $[HCO_3^-]$ in the range of 24–26 meq/L. Type 1 RTA patients typically respond to chronic alkali therapy readily, and the benefits of adequate alkali therapy include a decrease in the frequency of nephrolithiasis, improvement in bone density, resumption of normal growth patterns in children, and preservation of kidney function in both adults and children. For type 4 RTA, attention must be paid to the dual goals of correction of the metabolic acidosis, using the same approach as for classical distal renal tubular acidosis (type 1 RTA), and also correction of the plasma $[K^+]$. Restoration of normokalemia increases urinary net acid excretion and consequently can greatly improve the metabolic acidosis. Chronic administration of oral sodium polystyrene sulfonate (15 g of power prepared as an oral solution, without sorbitol, once daily 2–3 times per week) is sometimes used but is unpalatable, and patient compliance is low. The nonabsorbed, calcium-potassium cation exchange polymer, patiromer, may be considered for type 4 RTA patients with hyperkalemia because it is more palatable. It is administered as 8.4-g packets of powder for suspension PO twice daily with dose adjustment at weekly intervals, based on the hyperkalemic response. Hypoaldosteronism should receive fludrocortisone, but the dose varies with the cause of the hormone deficiency. This agent should be administered very cautiously and in combination with furosemide in patients with edema and hypertension because of possible aggravation of

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Non–Anion Gap Metabolic Acidoses

For non-AG acidosis due to gastrointestinal losses of bicarbonate, $NaHCO_3$ may be administered intravenously or orally, as determined by the severity of both the acidosis and the accompanying volume depletion. Proximal RTA is the most challenging of the RTAs to treat if the goal is to restore the serum [HCO_3^-] to normal because administration of oral alkali increases urinary excretion of bicarbonate and potassium. In patients with proximal RTA (type 2), potassium administration is typically required. An oral solution of sodium and potassium citrate (citric acid 334 mg, sodium citrate 500 mg, and potassium citrate 550 mg per 5 mL) may be prescribed for this purpose (Virtrate or Cytra-3). In classical distal RTA (type 1), hypokalemia should be corrected first. When accomplished, alkali therapy with either sodium citrate (Shohl's solution) or $NaHCO_3$ tablets (650-mg tablets contain 7.8 meq) should be initiated to correct and maintain the serum [HCO_3^-] in the range of 24–26 meq/L. Type 1 RTA patients typically respond to chronic alkali therapy readily, and the benefits of adequate alkali therapy include a decrease in the frequency of nephrolithiasis, improvement in bone density, resumption of normal growth patterns in children, and preservation of kidney function in both adults and children. For type 4 RTA, attention must be paid to the dual goals of correction of the metabolic acidosis, using the same approach as for classical distal renal tubular acidosis (type 1 RTA), and also correction of the plasma [K^+]. Restoration of normokalemia increases urinary net acid excretion and consequently can greatly improve the metabolic acidosis. Chronic administration of oral sodium polystyrene sulfonate (15 g of power prepared as an oral solution, without sorbitol, once daily 2–3 times per week) is sometimes used but is unpalatable, and patient compliance is low. The nonabsorbed, calcium-potassium cation exchange polymer, patiromer, may be considered for type 4 RTA patients with hyperkalemia because it is more palatable. It is administered as 8.4-g packets of powder for suspension PO twice daily with dose adjustment at weekly intervals, based on the plasma [K^+], not to exceed 25.2 g/d. Additionally, the diet should be low in potassium-containing foods or supplements (salt substitute), all potassium-retaining medications should be discontinued, and a loop diuretic may be administered. Finally, patients with documented isolated hypoaldosteronism should receive fludrocortisone, but the dose varies with the cause of the hormone deficiency. This agent should be administered very cautiously and in combination with furosemide in patients with edema and hypertension because of possible aggravation of these conditions.

METABOLIC ALKALOSIS

Metabolic alkalosis is established by an elevated arterial pH, an increase in the serum [HCO_3^-], and an increase in $Paco_2$ as a result of compensatory alveolar hypoventilation (Table 55-1). It is often accompanied by hypochloremia and hypokalemia. The elevation in arterial pH establishes the diagnosis because pH is decreased in respiratory acidosis, even though both have an elevated $Paco_2$. Metabolic alkalosis frequently occurs as a mixed acid-base disorder in association with either respiratory acidosis, respiratory alkalosis, or metabolic acidosis.

ETIOLOGY AND PATHOGENESIS

Metabolic alkalosis occurs as a result of net gain of [HCO_3^-] or loss of nonvolatile acid (usually HCl by vomiting) from the extracellular fluid. When vomiting causes loss of HCl from the stomach, HCO_3^- secretion cannot be initiated in the small bowel, and thus, HCO_3^- is retained in the extracellular fluid. Thus, vomiting or nasogastric suction is an example of the *generation stage* of metabolic alkalosis, in which the loss of acid typically causes alkalosis. Upon cessation of vomiting, the *maintenance stage* ensues because secondary factors prevent the kidneys from excreting HCO_3^- appropriately.

Maintenance of metabolic alkalosis, therefore, represents a failure of the kidneys to eliminate excess HCO_3^- from the extracellular compartment. The kidneys will retain, rather than excrete, the excess alkali and maintain the alkalosis if (1) volume deficiency, chloride deficiency, and K^+ deficiency exist in combination with a reduced GFR (associated with a low urine [Cl^-]) or (2) hypokalemia exists because of autonomous hyperaldosteronism (normal urine [Cl^-]). In the first example, saline-responsive metabolic alkalosis is corrected by extracellular fluid volume (ECFV) restoration (IV administration of NaCl and KCl), whereas, in the latter, it may be necessary to repair the alkalosis by pharmacologic or surgical intervention, not with saline

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

To establish the cause of metabolic alkalosis **(Table 55-6)**, it is necessary to assess the status of the ECFV, the recumbent and upright blood pressure

NaCl and KCl), whereas, in the latter, it may be necessary to repair the alkalosis by pharmacologic or surgical intervention, not with saline administration (saline-unresponsive metabolic alkalosis).

DIFFERENTIAL DIAGNOSIS

To establish the cause of metabolic alkalosis **(Table 55-6)**, it is necessary to assess the status of the ECFV, the recumbent and upright blood pressure (to determine if orthostasis is present), the serum $[K^+]$, the urine $[Cl^-]$, and in some circumstances, the renin-aldosterone system. For example, the presence of chronic hypertension and chronic hypokalemia in an alkalotic patient suggests either mineralocorticoid excess or that the hypertensive patient is receiving diuretics. Low plasma renin activity and values for urine $[Cl^-]$ >20 meq/L in a patient who is not taking diuretics suggest primary mineralocorticoid excess. The combination of hypokalemia and alkalosis in a normotensive, nonedematous patient can be due to Bartter's or Gitelman's syndrome, magnesium deficiency, vomiting, exogenous alkali, or diuretic ingestion. Measurement of urine electrolytes (especially the urine $[Cl^-]$) and screening of the urine for diuretics are recommended. If the urine is alkaline, with an elevated $[Na^+]_u$ and $[K^+]_u$ but low $[Cl^-]_u$, the diagnosis is usually either vomiting (overt or surreptitious) or alkali ingestion. If the urine is relatively acid with low concentrations of Na^+, K^+, and Cl^-, the most likely possibilities are prior vomiting, the posthypercapnic state, or prior diuretic ingestion. If the urine sodium, potassium, and chloride concentrations are not depressed, magnesium deficiency, Bartter's or Gitelman's syndrome, or current diuretic ingestion should be considered. Bartter's syndrome is distinguished from Gitelman's syndrome by the presence of hypocalciuria in the latter disorder.

TABLE 55-6

Causes of Metabolic Alkalosis

I. Exogenous HCO_3^- loads
   A. Acute alkali administration
   B. Milk-alkali syndrome

II. Effective ECFV contraction, normotension, K^+ deficiency, and secondary hyperreninemic hyperaldosteronism
   A. Gastrointestinal origin
      1. Vomiting
      2. Gastric aspiration
      3. Congenital chloridorrhea
      4. Gastrocystoplasty
      5. Villous adenoma
   B. Renal origin
      1. Diuretic use (thiazides and loop diuretics)
      2. Posthypercapnic state
      3. Hypercalcemia/hypoparathyroidism
      4. Recovery from lactic acidosis or ketoacidosis
      5. Nonreabsorbable anions including penicillin, carbenicillin
      6. Mg^{2+} deficiency
      7. K^+ depletion
      8. Bartter's syndrome (loss-of-function mutations of transporters and ion channels in TALH)
      9. Gitelman's syndrome (loss-of-function mutation of Na^+-Cl^- cotransporter in DCT)

III. ECFV expansion, hypertension, K^+ deficiency, and mineralocorticoid excess
   A. High renin
      1. Renal artery stenosis
      2. Accelerated hypertension
      3. Renin-secreting tumor
      4. Estrogen therapy
   B. Low renin
      1. Primary aldosteronism
         a. Adenoma
         a. 11β-Hydroxylase deficiency

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

I. Exogenous HCO_3^- loads
   A. Acute alkali administration
   B. Milk-alkali syndrome

II. Effective ECFV contraction, normotension, K^+ deficiency, and secondary hyperreninemic hyperaldosteronism
   A. Gastrointestinal origin
      1. Vomiting
      2. Gastric aspiration
      3. Congenital chloridorrhea
      4. Gastrocystoplasty
      5. Villous adenoma
   B. Renal origin
      1. Diuretic use (thiazides and loop diuretics)
      2. Posthypercapnic state
      3. Hypercalcemia/hypoparathyroidism
      4. Recovery from lactic acidosis or ketoacidosis
      5. Nonreabsorbable anions including penicillin, carbenicillin
      6. Mg^{2+} deficiency
      7. K^+ depletion
      8. Bartter's syndrome (loss-of-function mutations of transporters and ion channels in TALH)
      9. Gitelman's syndrome (loss-of-function mutation of Na^+-Cl^- cotransporter in DCT)

III. ECFV expansion, hypertension, K^+ deficiency, and mineralocorticoid excess
   A. High renin
      1. Renal artery stenosis
      2. Accelerated hypertension
      3. Renin-secreting tumor
      4. Estrogen therapy
   B. Low renin
      1. Primary aldosteronism
         a. Adenoma
         b. Hyperplasia
         c. Carcinoma
      2. Adrenal enzyme defects
         a. 11β-Hydroxylase deficiency
         b. 17α-Hydroxylase deficiency
      3. Cushing's syndrome or disease
      4. Other
         a. Licorice
         b. Carbenoxolone
         c. Chewer's tobacco

IV. Gain-of-function mutation of sodium channel in DCT with ECFV expansion, hypertension, K^+ deficiency, and hyporeninemic-hypoaldosteronism
   A. Liddle's syndrome

Abbreviations: DCT, distal convoluted tubule; ECFV, extracellular fluid volume; TALH, thick ascending limb of Henle's loop.

Alkali Administration

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

enhanced reabsorption of HCO_3^-. Such patients include those who receive $NaHCO_3$ (PO or IV), citrate loads IV (transfusions of whole blood, or therapeutic apheresis), or antacids plus cation-exchange resins (aluminum hydroxide and sodium polystyrene sulfonate). Nursing home patients

Chronic administration of alkali to individuals with normal renal function rarely causes alkalosis. However, in patients with coexistent hemodynamic disturbances associated with effective ECFV depletion (e.g., heart failure), alkalosis can develop because of diminished capacity to excrete HCO_3^- or enhanced reabsorption of HCO_3^-. Such patients include those who receive $NaHCO_3$ (PO or IV), citrate loads IV (transfusions of whole blood, or therapeutic apheresis), or antacids plus cation-exchange resins (aluminum hydroxide and sodium polystyrene sulfonate). Nursing home patients receiving enteral tube feedings have a higher incidence of metabolic alkalosis than nursing home patients receiving regular diets.

METABOLIC ALKALOSIS ASSOCIATED WITH ECFV CONTRACTION, K^+ DEPLETION, AND SECONDARY HYPERRENINEMIC HYPERALDOSTERONISM

Gastrointestinal Origin

Gastrointestinal loss of H^+ from vomiting or gastric aspiration causes simultaneous addition of HCO_3^- into the extracellular fluid. During active vomiting, the filtered load of bicarbonate reaching the kidneys is acutely increased and will exceed the reabsorptive capacity of the proximal tubule for HCO_3^- absorption. Subsequently, enhanced delivery of HCO_3^- to the distal nephron, where the capacity for HCO_3^- reabsorption is lower, will result in excretion of alkaline urine that stimulates potassium secretion. When vomiting ceases, the persistence of volume, potassium, and chloride depletion triggers maintenance of the alkalosis because these conditions promote HCO_3^- reabsorption. Correction of the contracted ECFV with NaCl and repair of K^+ deficits with KCl corrects the acid-base disorder by restoring the ability of the kidney to excrete the excess bicarbonate.

Renal Origin

DIURETICS

(See also Chap. 258) Diuretics such as thiazides and loop diuretics (furosemide, bumetanide, torsemide) increase excretion of salt and acutely diminish the ECFV without altering the total body bicarbonate content. The serum $[HCO_3^-]$ increases because the reduced ECFV "contracts" around the $[HCO_3^-]$ in plasma (contraction alkalosis). The chronic administration of diuretics tends to generate an alkalosis by increasing distal salt delivery so that both K^+ and H^+ secretion are stimulated. The alkalosis is maintained by persistence of the contraction of the ECFV, secondary hyperaldosteronism, K^+ deficiency, and the direct effect of the diuretic (as long as diuretic administration continues). Discontinuing the diuretic and providing isotonic saline to correct the ECFV deficit will repair the alkalosis.

SOLUTE LOSING DISORDERS: BARTTER'S SYNDROME AND GITELMAN'S SYNDROME

See Chap. 315.

NON-REABSORBABLE ANIONS AND MAGNESIUM DEFICIENCY

Administration of large quantities of the penicillin derivatives carbenicillin or ticarcillin cause their non-reabsorbable anions to appear in the distal tubule. This increases the transepithelial potential difference in the collecting tubule and thereby enhances H^+ and K^+ secretion. Mg^{2+} deficiency may occur with chronic administration of thiazide diuretics, alcoholism, and malnutrition, and in Gitelman's syndrome, it potentiates the development of hypokalemic alkalosis by enhancing distal acidification through stimulation of renin and hence aldosterone secretion.

POTASSIUM DEPLETION

Chronic K^+ depletion as a result of extreme dietary potassium insufficiency, diuretics, or alcohol abuse may initiate metabolic alkalosis by increasing urinary net acid excretion. The renal generation of NH_4^+ (ammoniagenesis) is upregulated directly by hypokalemia. Chronic K^+ deficiency also upregulates the H^+, K^+-ATPases in the distal tubule and collecting duct to increase K^+ absorption while simultaneously increasing H^+ secretion. Alkalosis associated with severe K^+ depletion is resistant to salt administration, but repair of the K^+ deficiency corrects the alkalosis. Potassium depletion often occurs concurrent with magnesium deficiency in alcoholics with malnutrition.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

When an underlying stimulus for the generation of lactic acid or ketoacid is corrected, such as shock or severe volume depletion by volume restoration, or with insulin therapy, the lactate or ketones are metabolized to yield an equivalent amount of HCO_3^-. Exogenous sources of HCO_3^- will be

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

Alkalosis associated with severe K^+ depletion is resistant to salt administration, but repair of the K^+ deficiency corrects the alkalosis. Potassium depletion often occurs concurrent with magnesium deficiency in alcoholics with malnutrition.

AFTER TREATMENT OF LACTIC ACIDOSIS OR KETOACIDOSIS

When an underlying stimulus for the generation of lactic acid or ketoacid is corrected, such as shock or severe volume depletion by volume restoration, or with insulin therapy, the lactate or ketones are metabolized to yield an equivalent amount of HCO_3^-. Exogenous sources of HCO_3^- will be additive to that amount generated by organic anion metabolism and may create a surfeit of HCO_3^- ("rebound alkalosis").

POSTHYPERCAPNIA

Prolonged CO_2 retention with chronic respiratory acidosis enhances renal HCO_3^- absorption and the generation of new HCO_3^- (increased net acid excretion). Metabolic alkalosis results from the persistently elevated [HCO_3^-] when the elevated $Paco_2$ is abruptly returned toward normal.

METABOLIC ALKALOSIS ASSOCIATED WITH ECFV EXPANSION, HYPERTENSION, AND MINERALOCORTICOID EXCESS

Increased aldosterone levels may be the result of autonomous primary adrenal overproduction or of secondary aldosterone release due to renal overproduction of renin. Mineralocorticoid excess increases net acid excretion and may result in metabolic alkalosis, which is typically exacerbated by associated K^+ deficiency. Salt retention and hypertension are due to upregulation of the epithelial Na^+ channel (ENaC) in the collecting tubule in response to aldosterone. The kaliuresis persists because of mineralocorticoid excess and stimulation of ENaC, causing an increase in transepithelial voltage, which enhances K^+ excretion. Persistent K^+ depletion may cause polydipsia and polyuria.

Liddle's syndrome **(Chap. 315)** results from an inherited gain-of-function mutation of genes that regulate the collecting duct Na^+ channel, ENaC. This rare monogenic form of hypertension is the result of volume expansion that secondarily suppresses aldosterone elaboration. Patients typically present with hypertension, hypokalemia, and metabolic alkalosis.

Symptoms

With metabolic alkalosis, changes in CNS and peripheral nervous system function are similar to those of hypocalcemia **(Chap. 409)**; symptoms include mental confusion; obtundation; and a predisposition to seizures, paresthesias, muscular cramping, tetany, aggravation of arrhythmias, and hypoxemia in chronic obstructive pulmonary disease. Related electrolyte abnormalities include hypokalemia and hypophosphatemia.

TREATMENT

Metabolic Alkalosis

The first goal of therapy is to correct the underlying stimulus for HCO_3^- generation. If primary aldosteronism or Cushing's syndrome is present, correction of the underlying cause will reverse the hypokalemia and alkalosis. [H^+] loss by the stomach or kidneys can be mitigated by the use of proton pump inhibitors or the discontinuation of diuretics. The second aspect of treatment is to eliminate factors that sustain the inappropriate increase in HCO_3^- reabsorption, such as ECFV contraction or K^+ deficiency. K^+ deficits should always be repaired. Isotonic saline is recommended to reverse the alkalosis when ECFV contraction is present. If associated conditions, such as congestive heart failure, preclude infusion of isotonic saline, renal HCO_3^- loss can be accelerated by administration of acetazolamide (125–250 mg IV), a carbonic anhydrase inhibitor, which is usually effective in patients with adequate renal function. However, acetazolamide triggers urinary K^+ losses and may cause hypokalemia that should be corrected. Dilute hydrochloric acid IV (0.1 *N* HCl) has been advocated in extreme cases of metabolic alkalosis but causes hemolysis and must be delivered slowly in a central vein. This preparation is not available generally and must be prepared in the pharmacy. Because serious errors or harm may occur, its use is not advised. Therapy in Liddle's syndrome should include a potassium-sparing diuretic (amiloride or triamterene) to inhibit ENaC and correct both the hypertension and the hypokalemia.

RESPIRATORY ACIDOSIS

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

recognized by an increase in $Paco_2$ and decrease in pH **(Table 55-7)**. In acute respiratory acidosis, there is a compensatory elevation in HCO_3^- (due to cellular buffering mechanisms) that increases 1 mmol/L for every 10-mmHg increase in $Paco_2$. In chronic respiratory acidosis (>24 h), renal adaptation

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

RESPIRATORY ACIDOSIS

Respiratory acidosis occurs as a result of severe pulmonary disease, respiratory muscle fatigue, or abnormalities in ventilatory control and is recognized by an increase in $Paco_2$ and decrease in pH **(Table 55-7)**. In acute respiratory acidosis, there is a compensatory elevation in HCO_3^- (due to cellular buffering mechanisms) that increases 1 mmol/L for every 10-mmHg increase in $Paco_2$. In chronic respiratory acidosis (>24 h), renal adaptation increases the [HCO_3^-] by 4 mmol/L for every 10-mmHg increase in $Paco_2$. The serum HCO_3^- usually does not increase above 38 mmol/L.

TABLE 55-7

Respiratory Acid-Base Disorders

I. Alkalosis
- A. Central nervous system stimulation
  1. Pain
  2. Anxiety, psychosis
  3. Fever
  4. Cerebrovascular accident
  5. Meningitis, encephalitis
  6. Tumor
  7. Trauma
- B. Hypoxemia or tissue hypoxia
  1. High altitude
  2. Pneumonia, pulmonary edema
  3. Aspiration
  4. Severe anemia
- C. Drugs or hormones
  1. Pregnancy, progesterone
  2. Salicylates
  3. Cardiac failure
- D. Stimulation of chest receptors
  1. Hemothorax
  2. Flail chest
  3. Cardiac failure
  4. Pulmonary embolism
- E. Miscellaneous
  1. Septicemia
  2. Hepatic failure
  3. Mechanical hyperventilation
  4. Heat exposure
  5. Recovery from metabolic acidosis

II. Acidosis
- A. Central
  1. Drugs (anesthetics, morphine, sedatives)
  2. Stroke
  3. Infection
- B. Airway
  1. Obstruction
  2. Asthma
- C. Parenchyma
  4. Adult respiratory distress syndrome
  5. Barotrauma

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

C. Parenchyma
   1. Emphysema
   2. Pneumoconiosis
   3. Bronchitis
   4. Adult respiratory distress syndrome
   5. Barotrauma

D. Neuromuscular
   1. Poliomyelitis
   2. Kyphoscoliosis
   3. Myasthenia
   4. Muscular dystrophies

E. Miscellaneous
   1. Obesity
   2. Hypoventilation
   3. Permissive hypercapnia

The clinical features vary according to the severity and duration of the respiratory acidosis, the underlying disease, and whether there is accompanying hypoxemia. A rapid increase in $Paco_2$ (acute hypercapnia) may cause anxiety, dyspnea, confusion, psychosis, and hallucinations and may progress to coma. However, chronic hypercapnia may cause sleep disorders; loss of memory; daytime somnolence; personality changes; impairment of coordination; and motor disturbances such as tremor, myoclonic jerks, and asterixis. Headaches and other signs that mimic raised intracranial pressure, such as papilledema, abnormal reflexes, and focal muscle weakness, are also seen.

Depression of the respiratory center by a variety of drugs, injury, or disease can produce respiratory acidosis. This may occur acutely with general anesthetics, sedatives, and head trauma or chronically with sedatives, alcohol, intracranial tumors, and the syndromes of sleep-disordered breathing including the primary alveolar and obesity-hypoventilation syndromes **(Chaps. 296 and 297)**. Abnormalities or disease in the motor neurons, neuromuscular junction, and skeletal muscle can cause hypoventilation via respiratory muscle fatigue. Mechanical ventilation, when not properly adjusted, may result in respiratory acidosis, particularly if CO_2 production suddenly rises (because of fever, agitation, sepsis, or overfeeding) or alveolar ventilation decreases because of worsening pulmonary function. High levels of positive end-expiratory pressure in the presence of reduced cardiac output may cause hypercapnia as a result of large increases in alveolar dead space **(Chap. 285)**. Permissive hypercapnia may be used to minimize intrinsic positive end-expiratory pressure in respiratory distress syndrome, but the consequential respiratory acidosis may require administration of $NaHCO_3$ to increase the arterial pH to approximately 7.20, but not to the normal value.

Acute hypercapnia follows sudden occlusion of the upper airway or generalized bronchospasm as in severe asthma, anaphylaxis, inhalational burn, or toxin injury. Chronic hypercapnia and respiratory acidosis occur in end-stage obstructive lung disease. Restrictive disorders involving both the chest wall and the lungs can cause respiratory acidosis because the high metabolic cost of respiration causes ventilatory muscle fatigue. Advanced stages of intrapulmonary and extrapulmonary restrictive defects present as chronic respiratory acidosis.

The diagnosis of respiratory acidosis requires the measurement of $Paco_2$ and arterial pH. A detailed history and physical examination often indicate the cause. Pulmonary function studies **(Chap. 285)**, including spirometry, diffusion capacity for carbon monoxide, lung volumes, and arterial $Paco_2$ and O_2 saturation, usually make it possible to determine if respiratory acidosis is secondary to lung disease. The workup for nonpulmonary causes should include a detailed drug history, measurement of hematocrit, and assessment of upper airway, chest wall, pleura, and neuromuscular function.

TREATMENT

Respiratory Acidosis

The management of respiratory acidosis depends on its severity and rate of onset. Acute respiratory acidosis can be life-threatening, and measures to reverse the underlying cause should be undertaken simultaneously with restoration of adequate alveolar ventilation. This may necessitate tracheal intubation and assisted mechanical ventilation. Oxygen administration should be titrated carefully in patients with severe obstructive

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

perfusion, and seizures). The $Paco_2$ should be lowered gradually in chronic respiratory acidosis, aiming to restore the $Paco_2$ to baseline levels and

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Respiratory Acidosis

The management of respiratory acidosis depends on its severity and rate of onset. Acute respiratory acidosis can be life-threatening, and measures to reverse the underlying cause should be undertaken simultaneously with restoration of adequate alveolar ventilation. This may necessitate tracheal intubation and assisted mechanical ventilation. Oxygen administration should be titrated carefully in patients with severe obstructive pulmonary disease and chronic CO_2 retention who are breathing spontaneously **(Chap. 292)**. When oxygen is used injudiciously, these patients may experience progression of the respiratory acidosis causing severe acidemia. Aggressive and rapid correction of hypercapnia should be avoided, because the falling $Paco_2$ may provoke the same complications noted with acute respiratory alkalosis (i.e., cardiac arrhythmias, reduced cerebral perfusion, and seizures). The $Paco_2$ should be lowered gradually in chronic respiratory acidosis, aiming to restore the $Paco_2$ to baseline levels and to provide sufficient Cl^- and K^+ to enhance the renal excretion of HCO_3^-.

Chronic respiratory acidosis is frequently difficult to correct, but the primary goal is to institute measures that may improve lung function **(Chap. 292)**.

RESPIRATORY ALKALOSIS

Alveolar hyperventilation decreases $Paco_2$ and increases the $HCO_3^-/Paco_2$ ratio, thus increasing pH (Table 55-7). Nonbicarbonate cellular buffers respond by consuming HCO_3^-. Hypocapnia develops when a sufficiently strong ventilatory stimulus causes CO_2 output in the lungs to exceed its metabolic production by tissues. Plasma pH and $[HCO_3^-]$ appear to vary proportionately with $Paco_2$ over a range from 40–15 mmHg. The relationship between arterial $[H^+]$ concentration and $Paco_2$ is ~0.7 mmol/L per mmHg (or 0.01 pH unit/mmHg), and that for plasma $[HCO_3^-]$ is 0.2 mmol/L per mmHg. Hypocapnia sustained for >2–6 h is further compensated by a decrease in renal ammonium and titratable acid excretion and a reduction in filtered HCO_3^- reabsorption. Full renal adaptation to respiratory alkalosis may take several days and requires normal volume status and renal function. The kidneys appear to respond directly to the lowered $Paco_2$ rather than to alkalosis per se. In chronic respiratory alkalosis, a 1-mmHg decrease in $Paco_2$ causes a 0.4-to 0.5-mmol/L drop in $[HCO_3^-]$ and a 0.3-mmol/L decrease in $[H^+]$ (or 0.003 increase in pH).

The effects of respiratory alkalosis vary according to duration and severity but are primarily those of the underlying disease. Reduced cerebral blood flow as a consequence of a rapid decline in $Paco_2$ may cause dizziness, mental confusion, and seizures, even in the absence of hypoxemia. The cardiovascular effects of acute hypocapnia in the conscious human are generally minimal, but in the anesthetized or mechanically ventilated patient, cardiac output and blood pressure may fall because of the depressant effects of anesthesia and positive-pressure ventilation on heart rate, systemic resistance, and venous return. Cardiac arrhythmias may occur in patients with heart disease as a result of changes in oxygen unloading by blood from a left shift in the hemoglobin-oxygen dissociation curve (Bohr effect). Acute respiratory alkalosis causes intracellular shifts of Na^+, K^+, and PO_4^{2-} and reduces free $[Ca^{2+}]$ by increasing the protein-bound fraction. Hypocapnia-induced hypokalemia is usually minor.

Chronic respiratory alkalosis is the most common acid-base disturbance in critically ill patients and, when severe, portends a poor prognosis. Many cardiopulmonary disorders manifest respiratory alkalosis in their early to intermediate stages, and the finding of normocapnia and hypoxemia in a patient with hyperventilation may herald the onset of rapid respiratory failure and should prompt an assessment to determine if the patient is becoming fatigued. Respiratory alkalosis is common during mechanical ventilation.

The hyperventilation syndrome may be disabling. Paresthesia; circumoral numbness; chest wall tightness or pain; dizziness; inability to take an adequate breath; and, rarely, tetany may be sufficiently stressful to perpetuate the disorder. Arterial blood-gas analysis demonstrates an acute or chronic respiratory alkalosis, often with hypocapnia in the range of 15–30 mmHg and no hypoxemia. CNS diseases or injury can produce several patterns of hyperventilation and sustained $Paco_2$ levels of 20–30 mmHg. Hyperthyroidism, high caloric loads, and exercise raise the basal metabolic rate, but ventilation usually rises in proportion so that arterial blood gases are unchanged and respiratory alkalosis does not develop. Salicylates are

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Respiratory alkalosis is also prominent in liver failure, and the severity correlates with the degree of hepatic insufficiency. Respiratory alkalosis is often

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

rate, but ventilation usually rises in proportion so that arterial blood gases are unchanged and respiratory alkalosis does not develop. Salicylates are the most common cause of drug-induced respiratory alkalosis because of direct stimulation of the medullary chemoreceptor **(Chap. 458)**. In addition, the methylxanthines, theophylline, and aminophylline stimulate ventilation and increase the ventilatory response to CO_2. Progesterone increases ventilation and lowers arterial $Paco_2$ by as much as 5–10 mmHg. Therefore, chronic respiratory alkalosis is a common feature of pregnancy. Respiratory alkalosis is also prominent in liver failure, and the severity correlates with the degree of hepatic insufficiency. Respiratory alkalosis is often an early finding in gram-negative septicemia, before fever, hypoxemia, or hypotension develops.

The diagnosis of respiratory alkalosis depends on measurement of arterial pH and $Paco_2$. The plasma $[K^+]$ is often reduced and the $[Cl^-]$ increased. In the acute phase, respiratory alkalosis is not associated with increased renal HCO_3^- excretion, but within hours, net acid excretion is reduced. In general, the HCO_3^- concentration falls by 2.0 mmol/L for each 10-mmHg decrease in $Paco_2$. Chronic respiratory alkalosis occurs when hypocapnia persists for greater than 3–5 days. The decline in $Paco_2$ reduces the serum $[HCO_3^-]$ by 4.0–5 mmol/L for each 10-mmHg decrease in $Paco_2$. It is unusual to observe a plasma HCO_3^- <12 mmol/L as a result of a pure respiratory alkalosis. The compensatory reduction in plasma $[HCO_3^-]$ is so effective in chronic respiratory alkalosis that the pH may not decline significantly from the normal value. Therefore, chronic respiratory alkalosis is the only acid-base disorder for which compensation can return the pH to the normal value.

When the diagnosis of respiratory alkalosis is made, its cause should be investigated. The diagnosis of hyperventilation syndrome is made by exclusion. In difficult cases, it may be important to rule out other conditions such as pulmonary embolism, coronary artery disease, and hyperthyroidism.

TREATMENT

Respiratory Alkalosis

The management of respiratory alkalosis is directed toward alleviation of the underlying disorder. If respiratory alkalosis complicates ventilator management, changes in dead space and tidal volume can minimize the hypocapnia. Patients with the hyperventilation syndrome may benefit from reassurance, rebreathing from a paper bag during symptomatic attacks, and attention to underlying psychological stress. Antidepressants and sedatives are not recommended. β-Adrenergic blockers may ameliorate peripheral manifestations of the hyperadrenergic state.

REFERENCES

Berend K et al: Physiological approach to assessment of acid-base disturbances. N Engl J Med 371:1434, 2014. [PubMed: 25295502]

Dubose TD: Etiologic causes of metabolic acidosis II: The normal anion gap acidosis. In *Metabolic Acidosis*. Wesson DE (ed). New York, Springer, 2016, pp 27–38.

Hamm LL, Dubose TD: Disorders of acid-base balance. In *Brenner and Rector's The Kidney,* 11th ed. Yu A et al (eds). Philadelphia, Elsevier, 2020, pp 496–536.

Kraut JA, Madias NE: Metabolic acidosis of CKD: An update. Am J Kidney Dis 67:307, 2016. [PubMed: 26477665]

Kraut JA, Madias NE: Re-evaluation of the normal range of serum total CO_2 concentration. Clin J Am Soc Nephrol 13:343, 2018. [PubMed: 29339355]

Palmer BF, Clegg DJ: Electrolyte and acid–base disturbances in patients with diabetes mellitus. N Engl J Med 373:548, 2015. [PubMed: 26244308]

Wesson DE et al: Mechanisms of metabolic acidosis-induced kidney injury in chronic kidney disease. J Am Soc Nephrol 31:469, 2020. [PubMed: 31988269]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 55: Acidosis and Alkalosis, Thomas D. DuBose, Jr.
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders

Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick

ECZEMA AND DERMATITIS

Eczema is a type of dermatitis, and these terms are often used synonymously (e.g., atopic eczema or atopic dermatitis [AD]). Eczema is a reaction pattern that presents with variable clinical findings and the common histologic finding of *spongiosis* (intercellular edema of the epidermis). Eczema is the final common expression for a number of disorders, including those discussed in the following sections. Primary lesions may include erythematous macules, papules, and vesicles, which can coalesce to form patches and plaques. In severe eczema, secondary lesions from infection or excoriation, marked by weeping and crusting, may predominate. In chronic eczematous conditions, *lichenification* (cutaneous hypertrophy and accentuation of normal skin markings) may alter the characteristic appearance of eczema.

ATOPIC DERMATITIS

AD is the cutaneous expression of the atopic state, characterized by a family history of asthma, allergic rhinitis, or eczema. The prevalence of AD is increasing worldwide. Some of its features are shown in **Table 57-1**. The etiology of AD is only partially defined, but there is a clear genetic predisposition. When both parents are affected by AD, >80% of their children manifest the disease. When only one parent is affected, the prevalence drops to slightly >50%. A characteristic defect in AD that contributes to the pathophysiology is an impaired epidermal barrier. In many patients, a mutation in the gene encoding filaggrin, a structural protein in the stratum corneum, is responsible. Patients with AD may display a variety of immunoregulatory abnormalities, including increased IgE synthesis; increased serum IgE levels; and impaired, delayed-type hypersensitivity reactions.

TABLE 57-1

Clinical Features of Atopic Dermatitis

1. Pruritus and scratching
2. Course marked by exacerbations and remissions
3. Lesions typical of eczematous dermatitis
4. Personal or family history of atopy (asthma, allergic rhinitis, food allergies, or eczema)
5. Clinical course lasting >6 weeks
6. Lichenification of skin
7. Presence of dry skin

The clinical presentation often varies with age. Half of patients with AD present within the first year of life, and 80% present by 5 years of age. About 80% ultimately coexpress allergic rhinitis or asthma. The infantile pattern is characterized by weeping inflammatory patches and crusted plaques on the face, neck, and extensor surfaces. The childhood and adolescent patterns are typified by dermatitis of flexural skin, particularly in the antecubital and popliteal fossae **(Fig. 57-1)**. AD may resolve spontaneously, but approximately 40% of all individuals affected as children will have dermatitis in adult life. The distribution of lesions in adults may be similar to those seen in childhood; however, adults frequently have localized disease manifesting as lichen simplex chronicus or hand eczema (see below). In patients with localized disease, AD may be suspected because of a typical personal or family history or the presence of cutaneous stigmata of AD such as perioral pallor, an extra fold of skin beneath the lower eyelid (Dennie-Morgan folds), increased palmar skin markings, and an increased incidence of cutaneous infections, particularly with *Staphylococcus aureus*. Regardless of other manifestations, pruritus is a prominent characteristic of AD in all age groups and is exacerbated by dry skin. Many of the cutaneous findings in

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Atopic dermatitis. Hyperpigmentation, lichenification, and scaling in the antecubital fossae are seen in this patient with atopic dermatitis. *(Courtesy*

ACCESS Medicine®

Access Provided by:
SILVERCHAIR

folds), increased palmar skin markings, and an increased incidence of cutaneous infections, particularly with *Staphylococcus aureus*. Regardless of other manifestations, pruritus is a prominent characteristic of AD in all age groups and is exacerbated by dry skin. Many of the cutaneous findings in affected patients, such as lichenification, are secondary to rubbing and scratching.

FIGURE 57-1

Atopic dermatitis. Hyperpigmentation, lichenification, and scaling in the antecubital fossae are seen in this patient with atopic dermatitis. *(Courtesy of Robert Swerlick, MD.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TREATMENT

Atopic Dermatitis

Therapy for AD should include avoidance of cutaneous irritants, adequate moisturization through the application of emollients, judicious use of topical anti-inflammatory agents, and prompt treatment of secondary infection. Patients should be instructed to bathe no more often than daily, using warm or cool water, and to use only mild bath soap. Immediately after bathing, while the skin is still moist, a topical anti-inflammatory agent in a cream or ointment base should be applied to areas of dermatitis, and all other skin areas should be lubricated with a moisturizer. Approximately 30 g of a topical agent is required to cover the entire body surface of an average adult.

Low- to mid-potency topical glucocorticoids are employed in most treatment regimens for AD. Skin atrophy and the potential for systemic absorption are constant concerns, especially with more potent agents. Low-potency topical glucocorticoids or nonglucocorticoid anti-inflammatory agents should be selected for use on the face and in intertriginous areas to minimize the risk of skin atrophy. Three nonglucocorticoid anti-inflammatory agents approved by the U.S. Food and Drug Administration (FDA) are available for topical use in AD: tacrolimus ointment, pimecrolimus cream, and crisaborole ointment. These agents do not cause skin atrophy, nor do they suppress the hypothalamic-pituitary-adrenal axis. The first two agents are topical calcineurin inhibitors (TCIs), whereas crisaborole is a phosphodiesterase-4 inhibitor. Concerns regarding the potential for lymphomas in patients treated with TCIs have largely been unfounded. Currently, all three agents are more costly than topical glucocorticoids. Barrier-repair products that attempt to restore the impaired epidermal barrier are also nonglucocorticoid agents and are gaining popularity in the treatment of AD.

Secondary infection of eczematous skin may lead to exacerbation of AD. Crusted and weeping skin lesions may be infected with *S. aureus*. When secondary infection is suspected, eczematous lesions should be cultured and patients treated with systemic antibiotics active against *S. aureus*. The initial use of penicillinase-resistant penicillins or cephalosporins is preferable. Dicloxacillin or cephalexin (250 mg qid for 7–10 days) is generally adequate for adults; however, antibiotic selection must be directed by culture results and clinical response. More than 50% of *S. aureus* isolates are now methicillin resistant in some communities. Current recommendations for the treatment of infection with these community-acquired methicillin-resistant *S. aureus* (CA-MRSA) strains in adults include trimethoprim-sulfamethoxazole (one double-strength tablet bid), minocycline (100 mg bid), doxycycline (100 mg bid), or clindamycin (300–450 mg qid). Duration of therapy should be 7–10 days. Inducible resistance may limit

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Control of pruritus is essential for treatment, as AD often represents "an itch that rashes." Antihistamines are most often used to control pruritus.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TREATMENT

Atopic Dermatitis

Therapy for AD should include avoidance of cutaneous irritants, adequate moisturization through the application of emollients, judicious use of topical anti-inflammatory agents, and prompt treatment of secondary infection. Patients should be instructed to bathe no more often than daily, using warm or cool water, and to use only mild bath soap. Immediately after bathing, while the skin is still moist, a topical anti-inflammatory agent in a cream or ointment base should be applied to areas of dermatitis, and all other skin areas should be lubricated with a moisturizer. Approximately 30 g of a topical agent is required to cover the entire body surface of an average adult.

Low- to mid-potency topical glucocorticoids are employed in most treatment regimens for AD. Skin atrophy and the potential for systemic absorption are constant concerns, especially with more potent agents. Low-potency topical glucocorticoids or nonglucocorticoid anti-inflammatory agents should be selected for use on the face and in intertriginous areas to minimize the risk of skin atrophy. Three nonglucocorticoid anti-inflammatory agents approved by the U.S. Food and Drug Administration (FDA) are available for topical use in AD: tacrolimus ointment, pimecrolimus cream, and crisaborole ointment. These agents do not cause skin atrophy, nor do they suppress the hypothalamic-pituitary-adrenal axis. The first two agents are topical calcineurin inhibitors (TCIs), whereas crisaborole is a phosphodiesterase-4 inhibitor. Concerns regarding the potential for lymphomas in patients treated with TCIs have largely been unfounded. Currently, all three agents are more costly than topical glucocorticoids. Barrier-repair products that attempt to restore the impaired epidermal barrier are also nonglucocorticoid agents and are gaining popularity in the treatment of AD.

Secondary infection of eczematous skin may lead to exacerbation of AD. Crusted and weeping skin lesions may be infected with *S. aureus*. When secondary infection is suspected, eczematous lesions should be cultured and patients treated with systemic antibiotics active against *S. aureus*. The initial use of penicillinase-resistant penicillins or cephalosporins is preferable. Dicloxacillin or cephalexin (250 mg qid for 7–10 days) is generally adequate for adults; however, antibiotic selection must be directed by culture results and clinical response. More than 50% of *S. aureus* isolates are now methicillin resistant in some communities. Current recommendations for the treatment of infection with these community-acquired methicillin-resistant *S. aureus* (CA-MRSA) strains in adults include trimethoprim-sulfamethoxazole (one double-strength tablet bid), minocycline (100 mg bid), doxycycline (100 mg bid), or clindamycin (300–450 mg qid). Duration of therapy should be 7–10 days. Inducible resistance may limit clindamycin's usefulness. Such resistance can be detected by the double-disk diffusion test, which should be ordered if the isolate is erythromycin resistant and clindamycin sensitive. As an adjunct, antibacterial washes or dilute sodium hypochlorite baths (0.005% bleach) and intermittent nasal mupirocin may be useful.

Control of pruritus is essential for treatment, as AD often represents "an itch that rashes." Antihistamines are most often used to control pruritus. Diphenhydramine (25 mg every 4–6 h), hydroxyzine (10–25 mg every 6 h), and doxepin (10–25 mg at bedtime) are useful primarily due to their sedating action. Higher doses of these agents may be required, but sedation can become bothersome. Patients need to be counseled about driving or operating heavy equipment after taking these medications. When used at bedtime, sedating antihistamines may improve the patient's sleep. Although they are effective in urticaria, nonsedating antihistamines and selective H_2 blockers are of little use in controlling the pruritus of AD.

Treatment with systemic glucocorticoids should be limited to severe exacerbations unresponsive to topical therapy. In the patient with chronic AD, therapy with systemic glucocorticoids will generally clear the skin only briefly, and cessation of the systemic therapy will invariably be accompanied by a return, if not a worsening, of the dermatitis. For chronic severe AD poorly responsive to standard topical regimens, systemic agents may be considered. Cyclosporine is approved for treatment of severe recalcitrant AD in some European countries. Monitoring of renal function and secondary infections is required. Dupilumab, an interleukin 4 receptor blocker, is FDA approved for use in patients 6 years of age and older and provides more targeted immunomodulation and a better safety profile than cyclosporine. Patients who do not respond to conventional therapies should be considered for patch testing to rule out allergic contact dermatitis (ACD). The role of dietary allergens in AD is controversial, and there is little evidence that they play any role outside of infancy, during which a small percentage of patients with AD may be affected by food allergens.

LICHEN SIMPLEX CHRONICUS

Lichen simplex chronicus may represent the end stage of a variety of pruritic and eczematous disorders, including AD. It consists of a circumscribed plaque or plaques of lichenified skin due to chronic scratching or rubbing. Common areas involved include the posterior nuchal region, dorsum of the feet, and ankles. Treatment of lichen simplex chronicus centers on breaking the cycle of chronic itching and scratching. High-potency topical

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

CONTACT DERMATITIS

Access Provided by:
SILVERCHAIR

plaque or plaques of lichenified skin due to chronic scratching or rubbing. Common areas involved include the posterior nuchal region, dorsum of the feet, and ankles. Treatment of lichen simplex chronicus centers on breaking the cycle of chronic itching and scratching. High-potency topical glucocorticoids are helpful in most cases, but, in recalcitrant cases, application of topical glucocorticoids under occlusion or intralesional injection of glucocorticoids may be required.

CONTACT DERMATITIS

Contact dermatitis is an inflammatory skin process caused by an exogenous agent or agents that directly or indirectly injure the skin. In *irritant* contact dermatitis (ICD), this injury is caused by an inherent characteristic of a compound—for example, a concentrated acid or base. Agents that cause *allergic* contact dermatitis (ACD) induce an antigen-specific immune response (e.g., poison ivy dermatitis). The clinical lesions of contact dermatitis may be acute (wet and edematous) or chronic (dry, thickened, and scaly), depending on the persistence of the insult **(see Chap. 56, Fig. 56-10)**.

Irritant Contact Dermatitis

ICD is generally well demarcated and often localized to areas of thin skin (eyelids, intertriginous areas) or areas where the irritant was occluded. Lesions may range from minimal skin erythema to areas of marked edema, vesicles, and ulcers. Prior exposure to the offending agent is not necessary, and the reaction develops in minutes to a few hours. Chronic low-grade irritant dermatitis is the most common type of ICD, and the most common area of involvement is the hands (see below). The most common irritants encountered are chronic wet work, soaps, and detergents. Treatment should be directed toward the avoidance of irritants and the use of protective gloves or clothing.

Allergic Contact Dermatitis

ACD is a manifestation of delayed-type hypersensitivity mediated by memory T lymphocytes in the skin. Prior exposure to the offending agent is necessary to develop the hypersensitivity reaction, which may take as little as 12 h or as long as 72 h to develop. The most common cause of ACD is exposure to plants, especially to members of the family Anacardiaceae, including the genus *Toxicodendron.* Poison ivy, poison oak, and poison sumac are members of this genus and cause an allergic reaction marked by erythema, vesiculation, and severe pruritus. The eruption is often linear or angular, corresponding to areas where plants have touched the skin. The sensitizing antigen common to these plants is urushiol, an oleoresin containing the active ingredient pentadecylcatechol. The oleoresin may adhere to skin, clothing, tools, and pets, and contaminated articles may cause dermatitis even after prolonged storage. Blister fluid does not contain urushiol and is not capable of inducing skin eruption in exposed subjects.

TREATMENT

Contact Dermatitis

If contact dermatitis is suspected and an offending agent is identified and removed, the eruption will resolve. Usually, treatment with high-potency topical glucocorticoids is enough to relieve symptoms while the dermatitis runs its course. For patients who require systemic therapy, daily oral prednisone—beginning at 1 mg/kg, but usually ≤60 mg/d—is sufficient. The dose should be tapered over 2–3 weeks, and each daily dose should be taken in the morning with food.

Identification of a contact allergen can be a difficult and time-consuming task. ACD should be suspected in patients with dermatitis unresponsive to conventional therapy or with an unusual and patterned distribution. Patients should be questioned carefully regarding occupational exposures and topical medications. Common sensitizers include preservatives in topical preparations, nickel sulfate, potassium dichromate, thimerosal, neomycin sulfate, fragrances, formaldehyde, and rubber-curing agents. Patch testing is helpful in identifying these agents but should not be attempted when patients have widespread active dermatitis or are taking systemic glucocorticoids.

HAND ECZEMA

Hand eczema is a very common, chronic skin disorder in which both exogenous and endogenous factors play important roles. It may be associated with other cutaneous disorders such as AD, and contact with various agents may be involved. Hand eczema represents a large proportion of cases of occupation-associated skin disease. Chronic, excessive exposure to water and detergents, harsh chemicals, or allergens may initiate or aggravate this disorder. It may present with dryness and cracking of the skin of the hands as well as with variable amounts of erythema and edema. Often, the dermatitis will begin under rings, where water and irritants are trapped. *Dyshidrotic* eczema, a variant of hand eczema, presents with multiple,

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

disorder. It may present with dryness and cracking of the skin of the hands as well as with variable amounts of erythema and edema. Often, the dermatitis will begin under rings, where water and irritants are trapped. *Dyshidrotic* eczema, a variant of hand eczema, presents with multiple, intensely pruritic, small papules and vesicles on the thenar and hypothenar eminences and the sides of the fingers **(Fig. 57-2)**. Lesions tend to occur in crops that slowly form crusts and then heal.

FIGURE 57-2

Dyshidrotic eczema. This example is characterized by deep-seated vesicles and scaling on palms and lateral fingers, and the disease is often associated with an atopic diathesis. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The evaluation of a patient with hand eczema should include an assessment of potential occupation-associated exposures. The history should be directed to identifying possible irritant or allergen exposures.

TREATMENT

Hand Eczema

Therapy for hand eczema is directed toward avoidance of irritants, identification of possible contact allergens, treatment of coexistent infection, and application of topical glucocorticoids. Whenever possible, the hands should be protected by gloves, preferably vinyl. The use of rubber gloves (latex) to protect dermatitic skin is sometimes associated with the development of hypersensitivity reactions to components of the gloves, which could be either a type I hypersensitivity reaction to the latex (manifested by the development of hives, itching, angioedema, and possibly anaphylaxis within minutes to hours of exposure) or a type IV hypersensitivity reaction to rubber accelerators (with worsening of eczematous eruptions days after exposure). Patients can be treated with cool moist compresses followed by application of a mid- to high-potency topical glucocorticoid in a cream or ointment base. As in AD, treatment of secondary infection is essential for good control. In addition, patients with hand eczema should be examined for dermatophyte infection by potassium hydroxide (KOH) preparation and culture (see below).

NUMMULAR ECZEMA

Nummular eczema is characterized by circular or oval "coinlike" lesions, beginning as small edematous papules that become crusted and scaly. The etiology of nummular eczema is unknown, but dry skin is a contributing factor. Common locations are the trunk or the extensor surfaces of the extremities, particularly on the pretibial areas or dorsum of the hands. Nummular eczema occurs more frequently in men and is most common in middle age. The treatment of nummular eczema is similar to that for AD.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

middle age. The treatment of nummular eczema is similar to that for AD.

ASTEATOTIC ECZEMA

Asteatotic eczema, also known as *xerotic eczema* or "winter itch," is a mildly inflammatory dermatitis that develops in areas of extremely dry skin, especially during the dry winter months. Clinically, there may be considerable overlap with nummular eczema. This form of eczema accounts for many physician visits because of the associated pruritus. Fine cracks and scale, with or without erythema, characteristically develop in areas of dry skin, especially on the anterior surfaces of the lower extremities in elderly patients. Asteatotic eczema responds well to topical moisturizers and the avoidance of cutaneous irritants. Overbathing and the use of harsh soaps exacerbate asteatotic eczema.

STASIS DERMATITIS AND STASIS ULCERATION

Stasis dermatitis develops on the lower extremities secondary to venous incompetence and chronic edema. Patients may give a history of deep venous thrombosis and may have evidence of vein removal or varicose veins. Early findings in stasis dermatitis consist of mild erythema and scaling associated with pruritus. The typical initial site of involvement is the medial aspect of the ankle, often over a distended vein **(Fig. 57-3)**.

FIGURE 57-3

Stasis dermatitis. An example of stasis dermatitis showing erythematous, scaly, and oozing patches over the lower leg. Several stasis ulcers are also seen in this patient. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Stasis dermatitis may become acutely inflamed, with crusting and exudate. In this state, it is easily confused with cellulitis. Of note, symmetrical and bilateral involvement is more likely stasis dermatitis, whereas unilateral involvement may represent cellulitis. Chronic stasis dermatitis is often associated with dermal fibrosis that is recognized clinically as brawny edema of the skin. As the disorder progresses, the dermatitis becomes progressively pigmented due to chronic erythrocyte extravasation leading to cutaneous hemosiderin deposition. Stasis dermatitis may be complicated by secondary infection and contact dermatitis. Severe stasis dermatitis may precede the development of stasis ulcers.

TREATMENT

Stasis Dermatitis and Stasis Ulceration

Patients with stasis dermatitis and stasis ulceration benefit greatly from leg elevation and the routine use of compression stockings with a gradient of at least 30–40 mmHg. Stockings providing less compression, such as antiembolism hose, are poor substitutes. Use of emollients and/or mid-potency topical glucocorticoids and avoidance of irritants are also helpful in treating stasis dermatitis. Protection of the legs from injury, including scratching, and control of chronic edema are essential to prevent ulcers. Diuretics may be required to adequately control chronic edema.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

control itching, scratching, and additional trauma. Superficial bacterial cultures of chronic stasis ulcers often yield polymicrobial colonizers and are

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Stasis Dermatitis and Stasis Ulceration

Patients with stasis dermatitis and stasis ulceration benefit greatly from leg elevation and the routine use of compression stockings with a gradient of at least 30–40 mmHg. Stockings providing less compression, such as antiembolism hose, are poor substitutes. Use of emollients and/or mid-potency topical glucocorticoids and avoidance of irritants are also helpful in treating stasis dermatitis. Protection of the legs from injury, including scratching, and control of chronic edema are essential to prevent ulcers. Diuretics may be required to adequately control chronic edema.

Stasis ulcers are difficult to treat, and resolution is slow. It is extremely important to elevate the affected limb as much as possible. The ulcer should be kept clear of necrotic material by gentle debridement and covered with a semipermeable dressing and a compression dressing or compression stocking. Glucocorticoids should not be applied to ulcers, because they may retard healing; however, they may be applied to the surrounding skin to control itching, scratching, and additional trauma. Superficial bacterial cultures of chronic stasis ulcers often yield polymicrobial colonizers and are of little utility in determination of secondary infection. Care must be taken to exclude treatable causes of leg ulcers (hypercoagulation, vasculitis, arterial insufficiency) before beginning the chronic management outlined above.

SEBORRHEIC DERMATITIS

Seborrheic dermatitis is a common, chronic disorder characterized by greasy scales overlying erythematous patches or plaques. Induration and scale are generally less prominent than in psoriasis, but clinical overlap exists between these diseases (“sebopsoriasis”). The most common location is in the scalp, where it may be recognized as severe dandruff. On the face, seborrheic dermatitis affects the eyebrows, eyelids, glabella, and nasolabial folds **(Fig. 57-4)**. Scaling of the external auditory canal is common in seborrheic dermatitis. In addition, the postauricular areas often become macerated and tender. Seborrheic dermatitis may also develop in the central chest, axilla, groin, submammary folds, and gluteal cleft. Rarely, it may cause widespread generalized dermatitis. Pruritus is variable.

FIGURE 57-4

Seborrheic dermatitis. Central facial erythema with overlying greasy, yellowish scale is seen in this patient. *(Courtesy of Jean Bolognia, MD; with permission.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Seborrheic dermatitis may be evident within the first few weeks of life, and within this context, it typically occurs in the scalp (“cradle cap”), face, or groin. It is rarely seen in children beyond infancy but becomes evident again during adolescent and adult life. Although it is frequently seen in patients

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

groin. It is rarely seen in children beyond infancy but becomes evident again during adolescent and adult life. Although it is frequently seen in patients with Parkinson's disease, in those who have had cerebrovascular accidents, and in those with HIV infection, the overwhelming majority of individuals with seborrheic dermatitis have no underlying disorder.

TREATMENT

Seborrheic Dermatitis

Treatment with low-potency topical glucocorticoids in conjunction with a topical antifungal agent, such as ketoconazole cream or ciclopirox cream, is often effective. The scalp and beard areas may benefit from antidandruff shampoos, which should be left in place 3–5 min before rinsing. High-potency topical glucocorticoid solutions (betamethasone or clobetasol) are effective for control of severe scalp involvement. High-potency glucocorticoids should not be used on the face because this treatment is often associated with steroid-induced rosacea or atrophy.

PAPULOSQUAMOUS DISORDERS

(TABLE 57-2)

TABLE 57-2

Papulosquamous Disorders

	CLINICAL FEATURES	OTHER NOTABLE FEATURES	HISTOLOGIC FEATURES
Psoriasis	Sharply demarcated, erythematous plaques with mica-like scale; predominantly on elbows, knees, and scalp; atypical forms may localize to intertriginous areas; eruptive forms may be associated with infection	May be aggravated by certain drugs, infection; severe forms seen in association with HIV	Acanthosis, vascular proliferation
Lichen planus	Purple polygonal papules marked by severe pruritus; lacy white markings, especially associated with mucous membrane lesions	Certain drugs may induce: thiazides, antimalarial drugs	Interface dermatitis
Pityriasis rosea	Rash often preceded by herald patch; oval to round plaques with trailing scale; most often affects trunk; eruption lines up in skinfolds giving a "fir tree–like" appearance; generally spares palms and soles	Variable pruritus; self-limited, resolving in 2–8 weeks; may be imitated by secondary syphilis	Pathologic features often nonspecific
Dermatophytosis	Polymorphous appearance depending on dermatophyte, body site, and host response; sharply defined to ill-demarcated scaly plaques with or without inflammation; may be associated with hair loss	KOH preparation may show branching hyphae; culture helpful	Hyphae and neutrophils in stratum corneum

Abbreviations: HIV, human immunodeficiency virus; KOH, potassium hydroxide.

PSORIASIS

Psoriasis is one of the most common dermatologic diseases, affecting up to 2% of the world's population. It is an immune-mediated disease clinically characterized by erythematous, sharply demarcated papules and rounded plaques covered by silvery micaceous scale. The skin lesions of psoriasis are variably pruritic. Traumatized areas often develop lesions of psoriasis (the *Koebner* or isomorphic phenomenon). In addition, other external factors may exacerbate psoriasis, including infections, stress, and medications (lithium, beta blockers, and antimalarial drugs).

The most common variety of psoriasis is called *plaque-type*. Patients with plaque-type psoriasis have stable, slowly enlarging plaques, which remain basically unchanged for long periods of time. The most commonly involved areas are the elbows, knees, gluteal cleft, and scalp. Involvement tends to

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Guttate psoriasis (eruptive psoriasis) is most common in children and young adults. It develops acutely in individuals without psoriasis or in those with

characterized by erythematous, sharply demarcated papules and rounded plaques covered by silvery micaceous scale. The skin lesions of psoriasis

ACCESS Medicine

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

The most common variety of psoriasis is called *plaque-type*. Patients with plaque-type psoriasis have stable, slowly enlarging plaques, which remain basically unchanged for long periods of time. The most commonly involved areas are the elbows, knees, gluteal cleft, and scalp. Involvement tends to be symmetric. Plaque psoriasis generally develops slowly and runs an indolent course. It rarely remits spontaneously. *Inverse psoriasis* affects the intertriginous regions, including the axilla, groin, submammary region, and navel; it also tends to affect the scalp, palms, and soles. The individual lesions are sharply demarcated plaques **(see Chap. 56, Fig. 56-7)**, but they may be moist and without scale due to their locations.

Guttate psoriasis (eruptive psoriasis) is most common in children and young adults. It develops acutely in individuals without psoriasis or in those with chronic plaque psoriasis. Patients present with many small erythematous, scaling papules, frequently after upper respiratory tract infection with β-hemolytic streptococci. The differential diagnosis should include pityriasis rosea and secondary syphilis.

In *pustular psoriasis*, patients may have disease localized to the palms and soles, or the disease may be generalized. Regardless of the extent of disease, the skin is erythematous, with pustules and variable scale. Localized to the palms and soles, it is easily confused with dishydrotic eczema. When it is generalized, episodes are characterized by fever (39°–40°C [102.2°–104.0°F]) lasting several days, an accompanying generalized eruption of sterile pustules, and a background of intense erythema; patients may become erythrodermic. Episodes of fever and pustules are recurrent. Local irritants, pregnancy, medications, infections, and systemic glucocorticoid withdrawal can precipitate this form of psoriasis. Oral retinoids are the treatment of choice in nonpregnant patients.

Fingernail involvement, appearing as punctate pitting, onycholysis, nail thickening, or subungual hyperkeratosis, may be a clue to the diagnosis of psoriasis when the clinical presentation is not classic.

According to the National Psoriasis Foundation, up to 30% of patients with psoriasis have psoriatic arthritis (PsA). It develops most commonly between the ages of 30 and 50 years. There are five subtypes of PsA: symmetric PsA, asymmetric PsA, distal PsA, spondylitis, and arthritis mutilans. Approximately 50% of PsA is classified as symmetric, which may resemble rheumatoid arthritis. Asymmetric arthritis comprises about 35% of cases. It can involve any joint and may present as “sausage digits.” Distal PsA is the classic form; however, it occurs in only about 5% of patients with PsA. It can involve fingers and toes; fingernails and toenails are often dystrophic, including nail pitting. Spondylitis also occurs in ~5% of patients with PsA. Arthritis mutilans is severe and deforming and affects primarily the small joints of the hands and feet. It accounts for fewer than 5% of PsA cases.

An increased risk of metabolic syndrome, including increased morbidity and mortality from cardiovascular events, has been demonstrated in psoriasis patients. Appropriate screening tests should be performed. The etiology of psoriasis is still poorly understood, but there is clearly a genetic component to the disease. In various studies, 30–50% of patients with psoriasis report a positive family history. Psoriatic lesions contain infiltrates of activated T cells that are thought to elaborate cytokines responsible for keratinocyte hyperproliferation, which results in the characteristic clinical findings. Agents inhibiting T-cell activation, clonal expansion, or release of proinflammatory cytokines are often effective for the treatment of severe psoriasis (see below).

TREATMENT

Psoriasis

Treatment of psoriasis depends on the type, location, and extent of disease. All patients should be instructed to avoid excess drying or irritation of their skin and to maintain adequate cutaneous hydration. Most cases of localized, plaque-type psoriasis can be managed with mid-potency topical glucocorticoids, although their long-term use is often accompanied by loss of effectiveness (tachyphylaxis) and atrophy of the skin. A topical vitamin D analogue (calcipotriene) and a retinoid (tazarotene) are also efficacious in the treatment of limited psoriasis and have largely replaced other topical agents such as coal tar, salicylic acid, and anthralin.

Ultraviolet (UV) light, natural or artificial, is an effective therapy for many patients with widespread psoriasis. Ultraviolet B (UVB), narrowband UVB, and ultraviolet A (UVA) light with either oral or topical psoralens (PUVA) are used clinically. UV light’s immunosuppressive properties are thought to be responsible for its therapeutic activity in psoriasis. It is also mutagenic, potentially leading to an increased incidence of nonmelanoma and melanoma skin cancer. UV-light therapy is contraindicated in patients receiving cyclosporine and should be used with great care in all immunocompromised patients due to the increased risk of skin cancer.

Various systemic agents can be used for severe, widespread psoriatic disease **(Table 57-3)**. Oral glucocorticoids should not be used for the treatment of psoriasis due to the potential for development of life-threatening pustular psoriasis when therapy is discontinued. Methotrexate is an effective agent, especially in patients with PsA. The synthetic retinoid acitretin is useful, especially when immunosuppression must be avoided; however, teratogenicity limits its use. Apremilast inhibits phosphodiesterase type 4. It is approved for both psoriasis and PsA. It must be used

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

biologic agents with more selective immunosuppressive properties and better safety profiles **(Table 57-4)**. These biologic agents appear to be

component to the disease. In various studies, 30–50% of patients with psoriasis report a positive family history. Psoriatic lesions contain infiltrates of

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

TREATMENT

Psoriasis

Treatment of psoriasis depends on the type, location, and extent of disease. All patients should be instructed to avoid excess drying or irritation of their skin and to maintain adequate cutaneous hydration. Most cases of localized, plaque-type psoriasis can be managed with mid-potency topical glucocorticoids, although their long-term use is often accompanied by loss of effectiveness (tachyphylaxis) and atrophy of the skin. A topical vitamin D analogue (calcipotriene) and a retinoid (tazarotene) are also efficacious in the treatment of limited psoriasis and have largely replaced other topical agents such as coal tar, salicylic acid, and anthralin.

Ultraviolet (UV) light, natural or artificial, is an effective therapy for many patients with widespread psoriasis. Ultraviolet B (UVB), narrowband UVB, and ultraviolet A (UVA) light with either oral or topical psoralens (PUVA) are used clinically. UV light's immunosuppressive properties are thought to be responsible for its therapeutic activity in psoriasis. It is also mutagenic, potentially leading to an increased incidence of nonmelanoma and melanoma skin cancer. UV-light therapy is contraindicated in patients receiving cyclosporine and should be used with great care in all immunocompromised patients due to the increased risk of skin cancer.

Various systemic agents can be used for severe, widespread psoriatic disease **(Table 57-3)**. Oral glucocorticoids should not be used for the treatment of psoriasis due to the potential for development of life-threatening pustular psoriasis when therapy is discontinued. Methotrexate is an effective agent, especially in patients with PsA. The synthetic retinoid acitretin is useful, especially when immunosuppression must be avoided; however, teratogenicity limits its use. Apremilast inhibits phosphodiesterase type 4. It is approved for both psoriasis and PsA. It must be used cautiously in the presence of renal failure or depression.

The evidence implicating psoriasis as a T-cell–mediated disorder has directed therapeutic efforts to immunoregulation. Cyclosporine and other immunosuppressive agents can be very effective in the treatment of psoriasis, and much attention is currently directed toward the development of biologic agents with more selective immunosuppressive properties and better safety profiles **(Table 57-4)**. These biologic agents appear to be quite efficacious in treatment of psoriasis and are well tolerated; however, caution with certain patient comorbidities must be exercised. Use of tumor necrosis factor-α (TNF-α) inhibitors may worsen congestive heart failure (CHF), and they should be used with caution in patients at risk for or known to have CHF. Further, none of the immunosuppressive agents used in the treatment of psoriasis should be initiated if the patient has a severe infection (including tuberculosis, HIV, hepatitis B or C); patients on such therapy should be routinely screened for tuberculosis. There have been reports of progressive multifocal leukoencephalopathy and lupus erythematosus in association with treatment with the TNF-α inhibitors. Malignancies, including a risk or history of certain malignancies, may limit the use of these systemic agents. In general, immunosuppressive agents have also been linked to an increase risk of skin cancer, and patients receiving these agents should be monitored for the development of skin cancer.

TABLE 57-3

FDA-Approved Systemic Therapy for Psoriasis

AGENT	MEDICATION CLASS	ADMINISTRATION		ADVERSE EVENTS (SELECTED)
		ROUTE	FREQUENCY	
Methotrexate	Antimetabolite	Oral	Weekly[a]	Hepatotoxicity, pulmonary toxicity, pancytopenia, potential for increased malignancies, ulcerative stomatitis, nausea, diarrhea, teratogenicity
Acitretin	Retinoid	Oral	Daily	Teratogenicity, hepatotoxicity, hyperostosis, hyperlipidemia/pancreatitis, depression, ophthalmologic effects, pseudotumor cerebri
Cyclosporine	Calcineurin inhibitor	Oral	Twice daily	Renal dysfunction, hypertension, hyperkalemia, hyperuricemia, hypomagnesemia, hyperlipidemia, increased risk of malignancies

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

[a]Initial test dose is required. [b]Initial dose escalation is required.

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Malignancies, including a risk of history of certain malignancies, may limit the use of these systemic agents. In general, immunosuppressive agents

TABLE 57-3

FDA-Approved Systemic Therapy for Psoriasis

AGENT	MEDICATION CLASS	ADMINISTRATION		ADVERSE EVENTS (SELECTED)
		ROUTE	FREQUENCY	
Methotrexate	Antimetabolite	Oral	Weekly[a]	Hepatotoxicity, pulmonary toxicity, pancytopenia, potential for increased malignancies, ulcerative stomatitis, nausea, diarrhea, teratogenicity
Acitretin	Retinoid	Oral	Daily	Teratogenicity, hepatotoxicity, hyperostosis, hyperlipidemia/pancreatitis, depression, ophthalmologic effects, pseudotumor cerebri
Cyclosporine	Calcineurin inhibitor	Oral	Twice daily	Renal dysfunction, hypertension, hyperkalemia, hyperuricemia, hypomagnesemia, hyperlipidemia, increased risk of malignancies
Apremilast	Phosphodiesterase type 4 inhibitor	Oral	Twice daily[b]	Hypersensitivity reaction, depression, nausea, diarrhea, vomiting, dyspepsia, weight loss, headache, fatigue

[a]Initial test dose is required. [b]Initial dose escalation is required.

Abbreviation: FDA, Food and Drug Administration.

TABLE 57-4

FDA-Approved Biologics for Psoriasis or Psoriatic Arthritis

MECHANISM OF ACTION	AGENTS (INDICATION; ROUTE)	FREQUENCY	WARNINGS, SELECTED
Anti-TNF-α	Etanercept (Ps, PsA; SC) Adalimumab (Ps, PsA; SC) Certolizumab (Ps, PsA; SC) Infliximab (Ps, PsA; IV) Golimumab (PsA; SC)	Ranges from once or twice weekly[a] to every 8 weeks[a]	Serious infections, hepatotoxicity, CHF, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies
Anti-IL-12 and anti-IL-23	Ustekinumab (Ps, PsA; SC)	Every 12 weeks[a]	Serious infections, neurologic events, potential for increased malignancies
Anti-IL-23	Risankizumab (Ps; SC) Tildrakizumab (Ps; SC) Guselkumab (Ps;	Ranges from every 8–12 weeks[a]	Serious infections, headaches
	Ixekizumab (Ps;		

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 57-4

FDA-Approved Biologics for Psoriasis or Psoriatic Arthritis

MECHANISM OF ACTION	AGENTS (INDICATION; ROUTE)	FREQUENCY	WARNINGS, SELECTED
Anti-TNF-α	Etanercept (Ps, PsA; SC) Adalimumab (Ps, PsA; SC) Certolizumab (Ps, PsA; SC) Infliximab (Ps, PsA; IV) Golimumab (PsA; SC)	Ranges from once or twice weekly[a] to every 8 weeks[a]	Serious infections, hepatotoxicity, CHF, hematologic events, hypersensitivity reactions, neurologic events, potential for increased malignancies
Anti-IL-12 and anti-IL-23	Ustekinumab (Ps, PsA; SC)	Every 12 weeks[a]	Serious infections, neurologic events, potential for increased malignancies
Anti-IL-23	Risankizumab (Ps; SC) Tildrakizumab (Ps; SC) Guselkumab (Ps; SC)	Ranges from every 8–12 weeks[a]	Serious infections, headaches
Anti-IL-17	Secukinumab (Ps, PsA; SC) Ixekizumab (Ps; SC) Brodalumab (Ps; SC)	Ranges from every 2–4 weeks[a]	Serious infections, hypersensitivity reaction, inflammatory bowel disease

[a]Initial dose modifications required.

Abbreviations: CHF, congestive heart failure; IL, interleukin; IV, intravenous; Ps, psoriasis; PsA, psoriatic arthritis; SC, subcutaneous; TNF-α, tumor necrosis factor-α.

LICHEN PLANUS

Lichen planus (LP) is a papulosquamous disorder that may affect the skin, scalp, nails, and mucous membranes. The primary cutaneous lesions are pruritic, polygonal, flat-topped, violaceous papules. Close examination of the surface of these papules often reveals a network of gray lines (*Wickham's striae*). The skin lesions may occur anywhere but have a predilection for the wrists, shins, lower back, and genitalia **(Fig. 57-5)**. Involvement of the scalp (*lichen planopilaris*) may lead to scarring alopecia, and nail involvement may lead to permanent deformity or loss of fingernails and toenails. LP commonly involves mucous membranes, particularly the buccal mucosa, where it can present on a spectrum ranging from a mild, white, reticulate eruption of the mucosa to a severe, erosive stomatitis. Erosive stomatitis may persist for years and may be linked to an increased risk of oral squamous cell carcinoma. Cutaneous eruptions clinically resembling LP have been observed after administration of numerous drugs, including thiazide diuretics, gold, antimalarial agents, penicillamine, and phenothiazines, and in patients with skin lesions of chronic graft-versus-host disease. In addition, LP may

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

striae). The skin lesions may occur anywhere but have a predilection for the wrists, shins, lower back, and genitalia (Fig. 57-5). Involvement of the cell carcinoma. Cutaneous eruptions clinically resembling LP have been observed after administration of numerous drugs, including thiazide diuretics, gold, antimalarial agents, penicillamine, and phenothiazines, and in patients with skin lesions of chronic graft-versus-host disease. In addition, LP may be associated with hepatitis C infection. The course of LP is variable, but most patients have spontaneous remissions 6 months to 2 years after the onset of disease. Topical glucocorticoids are the mainstay of therapy.

FIGURE 57-5

Lichen planus. An example of lichen planus showing multiple flat-topped, violaceous papules and plaques. Nail dystrophy, as seen in this patient's thumbnail, may also be a feature. *(Courtesy of Robert Swerlick, MD; with permission.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

PITYRIASIS ROSEA

Pityriasis rosea (PR) is a papulosquamous eruption of unknown etiology occurring more commonly in the spring and fall. Its first manifestation is the development of a 2- to 6-cm annular lesion (the herald patch). This is followed in a few days to a few weeks by the appearance of many smaller annular or papular lesions with a predilection to occur on the trunk (Fig. 57-6). The lesions are generally oval, with their long axis parallel to the skinfold lines. Individual lesions may range in color from red to brown and have a trailing scale. PR shares many clinical features with the eruption of secondary syphilis, but palm and sole lesions are extremely rare in PR and common in secondary syphilis. The eruption tends to be moderately pruritic and lasts 3–8 weeks. Treatment is directed at alleviating pruritus and consists of oral antihistamines; mid-potency topical glucocorticoids; and, in some cases, UVB phototherapy.

FIGURE 57-6

Pityriasis rosea. In this patient with pityriasis rosea, multiple round to oval erythematous patches with fine central scale are distributed along the skin tension lines on the trunk. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

CUTANEOUS INFECTIONS

(TABLE 57-5)

TABLE 57-5

Common Skin Infections

	CLINICAL FEATURES	ETIOLOGIC AGENT	TREATMENT
Impetigo	Honey-colored crusted papules, plaques, or bullae	Group A *Streptococcus* and *Staphylococcus aureus*	Systemic or topical antistaphylococcal and antistreptococcal antibiotics
Dermatophytosis	Inflammatory or noninflammatory annular scaly plaques; may involve hair loss; groin involvement spares scrotum; hyphae on KOH preparation	*Trichophyton, Epidermophyton*, or *Microsporum* spp.	Topical azoles, systemic griseofulvin, terbinafine, or azoles
Candidiasis	Inflammatory papules and plaques with satellite pustules, frequently in intertriginous areas; may involve scrotum; pseudohyphae on KOH preparation	*Candida albicans* and other *Candida* spp.	Topical nystatin or azoles; systemic azoles for resistant disease
Tinea versicolor	Hyper- or hypopigmented scaly patches on trunk; characteristic mixture of hyphae and spores ("spaghetti and meatballs") on KOH preparation	*Malassezia furfur*	Topical selenium sulfide lotion or azoles

Abbreviation: KOH, potassium hydroxide.

IMPETIGO, ECTHYMA, AND FURUNCULOSIS

Impetigo is a common superficial bacterial infection of skin caused most often by *S. aureus* **(Chap. 147)** and in some cases by group A β-hemolytic streptococci **(Chap. 148)**. The primary lesion is a superficial pustule that ruptures and forms a characteristic yellow-brown honey-colored crust **(see**

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

however, it should be considered along with toxic epidermal necrolysis and severe drug eruptions in patients with widespread blistering of the skin.

Abbreviation: KOH, potassium hydroxide.

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Impetigo is a common superficial bacterial infection of skin caused most often by *S. aureus* **(Chap. 147)** and in some cases by group A β-hemolytic streptococci **(Chap. 148)**. The primary lesion is a superficial pustule that ruptures and forms a characteristic yellow-brown honey-colored crust **(see Chap. 148, Fig. 148-3)**. Lesions may occur on normal skin (primary infection) or in areas already affected by another skin disease (secondary infection). Lesions caused by staphylococci may be tense, clear bullae, and this less common form of the disease is called *bullous impetigo*. Blisters are caused by the production of exfoliative toxin by *S. aureus* phage type II. This is the same toxin responsible for staphylococcal scalded-skin syndrome, often resulting in dramatic loss of the superficial epidermis due to blistering. The latter syndrome is much more common in children than in adults; however, it should be considered along with toxic epidermal necrolysis and severe drug eruptions in patients with widespread blistering of the skin. *Ecthyma* is a deep nonbullous variant of impetigo that causes punched-out ulcerative lesions. It is more often caused by a primary or secondary infection with *Streptococcus pyogenes*. Ecthyma is a deeper infection than typical impetigo and resolves with scars. Treatment of both ecthyma and impetigo involves gentle debridement of adherent crusts, facilitated by using soaks and topical antibiotics in conjunction with appropriate oral antibiotics.

Furunculosis is also caused by *S. aureus*, and this disorder has gained prominence in the past few decades because of CA-MRSA. A furuncle, or boil, is a painful, erythematous nodule that can occur on any cutaneous surface. The lesions may be solitary but are most often multiple. Patients frequently believe they have been bitten by spiders or insects. Family members or close contacts may also be affected. Furuncles can rupture and drain spontaneously or may need incision and drainage, which may be adequate therapy for small solitary furuncles without cellulitis or systemic symptoms. Whenever possible, lesional material should be sent for culture. Current recommendations for methicillin-sensitive infections are β-lactam antibiotics. Therapy for CA-MRSA is discussed previously (see "Atopic Dermatitis"). Warm compresses and nasal mupirocin are helpful therapeutic additions. Severe infections may require IV antibiotics.

ERYSIPELAS AND CELLULITIS

See Chap. 129.

DERMATOPHYTOSIS

Dermatophytes are fungi that infect skin, hair, and nails and include members of the genera *Trichophyton*, *Microsporum*, and *Epidermophyton* **(Chap. 219)**. *Tinea corporis*, or infection of the relatively hairless skin of the body (glabrous skin), may have a variable appearance depending on the extent of the associated inflammatory reaction. Typical infections consist of erythematous, scaly plaques, with an annular appearance that accounts for the common name "ringworm." Deep inflammatory nodules or granulomas occur in some infections, most often those inappropriately treated with mid- to high-potency topical glucocorticoids. Involvement of the groin (*tinea cruris*) is more common in males than in females. It presents as a scaling, erythematous eruption sparing the scrotum. Infection of the foot (*tinea pedis*) is the most common dermatophyte infection and is often chronic; it is characterized by variable erythema, edema, scaling, pruritus, and occasionally vesiculation. The infection may be widespread or localized but generally involves the web space between the fourth and fifth toes. Infection of the nails (*tinea unguium* or *onychomycosis*) occurs in many patients with tinea pedis and is characterized by opacified, thickened nails and subungual debris. The distal-lateral variant is most common. Proximal subungual onychomycosis may be a marker for HIV infection or other immunocompromised states. Dermatophyte infection of the scalp (*tinea capitis*) continues to be common, particularly affecting inner-city children but also affecting immunocompromised adults. The predominant organism is *Trichophyton tonsurans*, which can produce a relatively noninflammatory infection with mild scale and hair loss that is diffuse or localized. *T. tonsurans* and *Microsporum canis* can also cause a markedly inflammatory dermatosis with edema and nodules. This latter presentation is a *kerion*.

The diagnosis of tinea can be made from skin scrapings, nail scrapings, or hair by culture or direct microscopic examination with KOH. Nail clippings may be sent for histologic examination with periodic acid–Schiff (PAS) stain.

TREATMENT

Dermatophytosis

Both topical and systemic therapies may be used in dermatophyte infections. Treatment depends on the site involved and the type of infection. Topical therapy is generally effective for uncomplicated tinea corporis, tinea cruris, and limited tinea pedis. Topical agents are not effective as monotherapy for tinea capitis or onychomycosis (see below), and nystatin is not active against dermatophytes. Topicals are generally applied twice daily, and treatment should continue for 1 week beyond clinical resolution of the infection. Tinea pedis often requires longer treatment courses and frequently relapses. Oral antifungal agents may be required for recalcitrant tinea pedis or tinea corporis.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

are pregnant or breast-feeding.

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Dermatophytosis

Both topical and systemic therapies may be used in dermatophyte infections. Treatment depends on the site involved and the type of infection. Topical therapy is generally effective for uncomplicated tinea corporis, tinea cruris, and limited tinea pedis. Topical agents are not effective as monotherapy for tinea capitis or onychomycosis (see below), and nystatin is not active against dermatophytes. Topicals are generally applied twice daily, and treatment should continue for 1 week beyond clinical resolution of the infection. Tinea pedis often requires longer treatment courses and frequently relapses. Oral antifungal agents may be required for recalcitrant tinea pedis or tinea corporis.

For dermatophyte infections involving the hair and nails and for other infections unresponsive to topical therapy, oral antifungal agents are often used. Markedly inflammatory tinea capitis may result in scarring and hair loss, and a systemic antifungal agent plus systemic or topical glucocorticoids may be helpful in preventing these sequelae. A fungal etiology should be confirmed by direct microscopic examination or by culture before oral antifungal agents are prescribed for any infection. All the oral agents may cause hepatotoxicity. They should not be used in women who are pregnant or breast-feeding.

Griseofulvin is approved in the United States for dermatophyte infections involving the skin, hair, or nails. Common side effects of griseofulvin include gastrointestinal distress, headache, and urticaria.

Two other oral antifungal agents, itraconazole and terbinafine, are sometimes prescribed “off-label” for superficial fungal infections. Oral itraconazole is approved for onychomycosis. Itraconazole has the potential for serious interactions with other drugs requiring the P450 enzyme system for metabolism. Itraconazole should not be administered to patients with evidence of ventricular dysfunction or patients with known CHF.

Terbinafine is also approved for onychomycosis, and the granule version is approved for treatment of tinea capitis. Terbinafine has fewer interactions with other drugs than itraconazole; however, caution should be used with patients who are on multiple medications. The risk/benefit ratio should be considered when an asymptomatic toenail infection is treated with systemic agents.

The FDA has limited the use of a third oral agent due to potential hepatotoxicity and published the following: “Nizoral [ketoconazole] oral tablets should not be a first-line treatment for any fungal infection.” The topical form of ketoconazole is not affected by this action.

TINEA (PITYRIASIS) VERSICOLOR

Tinea versicolor is caused by a nondermatophytic, dimorphic fungus, *Malassezia furfur*, a normal inhabitant of the skin. The expression of infection is promoted by heat and humidity. The typical lesions consist of oval scaly macules, papules, and patches concentrated on the chest, shoulders, and back but only rarely on the face or distal extremities. On dark skin, the lesions often appear as hypopigmented areas, whereas on light skin, they are slightly erythematous or hyperpigmented. A KOH preparation from scaling lesions will demonstrate a confluence of short hyphae and round spores (“spaghetti and meatballs”). Lotions or shampoos containing sulfur, salicylic acid, or selenium sulfide are the treatments of choice and will clear the infection if used daily for 1–2 weeks and then weekly thereafter. These preparations are irritating if left on the skin for >10 min; thus, they should be washed off completely. Treatment with some oral antifungal agents is also effective, but they do not provide lasting results and are not FDA approved for this indication.

CANDIDIASIS

Candidiasis is a fungal infection caused by a related group of yeasts whose manifestations may be localized to the skin and mucous membranes or, rarely, may be systemic and life-threatening **(Chap. 216)**. The causative organism is usually *Candida albicans*. These organisms are normal saprophytic inhabitants of the gastrointestinal tract but may overgrow due to broad-spectrum antibiotic therapy, diabetes mellitus, or immunosuppression and cause disease. Candidiasis is a very common infection in HIV-infected individuals **(Chap. 202)**. The oral cavity is commonly involved. Lesions may occur on the tongue or buccal mucosa (*thrush*) and appear as white plaques. Fissured, macerated lesions at the corners of the mouth (*perlèche*) are often seen in individuals with poorly fitting dentures and may also be associated with candidal infection. In addition, candidal infections have an affinity for sites that are chronically wet and macerated, including the skin around nails (onycholysis and paronychia), and in intertriginous areas. Intertriginous lesions are characteristically edematous, erythematous, and scaly, with scattered “satellite pustules.” In males,

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TREATMENT

immunosuppression and cause disease. Candidiasis is a very common infection in HIV-infected individuals (Chap. 202). The oral cavity is commonly intertriginous areas. Intertriginous lesions are characteristically edematous, erythematous, and scaly, with scattered "satellite pustules." In males, there is often involvement of the penis and scrotum as well as the inner aspect of the thighs. In contrast to dermatophyte infections, candidal infections are frequently painful and accompanied by a marked inflammatory response. Diagnosis of candidal infection is based on the clinical pattern and demonstration of yeast on KOH preparation or culture.

TREATMENT

Candidiasis

Treatment involves removal of any predisposing factors such as antibiotic therapy or chronic moisture and the use of appropriate topical or systemic antifungal agents. Effective topicals include nystatin or azoles (miconazole, clotrimazole, econazole, or ketoconazole). The associated inflammatory response accompanying candidal infection on glabrous skin can be treated with a mild glucocorticoid lotion or cream (2.5% hydrocortisone). Systemic therapy is usually reserved for immunosuppressed patients or individuals with chronic or recurrent disease who fail to respond to appropriate topical therapy. Oral fluconazole is most commonly prescribed for cutaneous candidiasis. Oral nystatin is effective only for candidiasis of the gastrointestinal tract.

WARTS

Warts are cutaneous neoplasms caused by papillomaviruses. More than 100 different human papillomaviruses (HPVs) have been described. A typical wart, *verruca vulgaris*, is sessile, dome-shaped, and usually about a centimeter in diameter. Its surface is hyperkeratotic, consisting of many small filamentous projections. HPV also causes typical plantar warts, flat warts (*verruca plana*), and filiform warts. Plantar warts are endophytic and are covered by thick keratin. Paring of the wart will generally reveal a central core of keratinized debris and punctate bleeding points. Filiform warts are commonly seen on the face, neck, and skinfolds and present as papillomatous lesions on a narrow base. Flat warts are only slightly elevated and have a velvety, nonverrucous surface. They have a propensity for the face, arms, and legs, and are often spread by shaving.

Genital warts begin as small papillomas that may grow to form large, fungating lesions. In women, they may involve the labia, perineum, or perianal skin. In addition, the mucosa of the vagina, urethra, and anus can be involved as well as the cervical epithelium. In men, the lesions often occur initially in the coronal sulcus but may be seen on the shaft of the penis, the scrotum, or the perianal skin or in the urethra.

Appreciable evidence has accumulated indicating that HPV plays a role in the development of neoplasia of the uterine cervix and anogenital skin (Chap. 89). HPV types 16 and 18 have been most intensely studied and are the major risk factors for intraepithelial neoplasia and squamous cell carcinoma of the cervix, anus, vulva, and penis. The risk is higher among patients immunosuppressed after solid organ transplantation and among those infected with HIV. Recent evidence also implicates other HPV types. Histologic examination of biopsied samples from affected sites may reveal changes associated with typical warts and/or features typical of intraepidermal carcinoma (Bowen's disease). Squamous cell carcinomas associated with HPV infections have also been observed in extragenital skin (Chap. 76), most commonly in patients immunosuppressed after organ transplantation. Patients on long-term immunosuppression should be monitored for the development of squamous cell carcinoma and other cutaneous malignancies.

TREATMENT

Warts

Treatment of warts, other than anogenital warts, should be tempered by the observation that most warts in normal individuals resolve spontaneously within 1–2 years. There are many modalities available to treat warts, but no single therapy is universally effective. Factors that influence the choice of therapy include the location of the wart, the extent of disease, the age and immunologic status of the patient, and the patient's desire for therapy. Perhaps the most useful and convenient method for treating warts in almost any location is cryotherapy with liquid nitrogen. Equally effective for nongenital warts, but requiring much more patient compliance, is the use of keratolytic agents such as salicylic acid plasters or solutions. For genital warts, in-office application of a podophyllin solution is moderately effective but may be associated with marked local reactions. Prescription preparations of dilute, purified podophyllin are available for home use. Topical imiquimod, a potent inducer of local cytokine release, has been approved for treatment of genital warts. A topical compound composed of green tea extracts (sinecatechins) is also available. Conventional and laser surgical procedures may be required for recalcitrant warts. Recurrence of warts appears to be common with all these modalities. A highly effective vaccine for selected types of HPV has been approved by the FDA, and its use is reported to reduce the incidence

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

HERPES SIMPLEX

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TREATMENT

Warts

Treatment of warts, other than anogenital warts, should be tempered by the observation that most warts in normal individuals resolve spontaneously within 1–2 years. There are many modalities available to treat warts, but no single therapy is universally effective. Factors that influence the choice of therapy include the location of the wart, the extent of disease, the age and immunologic status of the patient, and the patient's desire for therapy. Perhaps the most useful and convenient method for treating warts in almost any location is cryotherapy with liquid nitrogen. Equally effective for nongenital warts, but requiring much more patient compliance, is the use of keratolytic agents such as salicylic acid plasters or solutions. For genital warts, in-office application of a podophyllin solution is moderately effective but may be associated with marked local reactions. Prescription preparations of dilute, purified podophyllin are available for home use. Topical imiquimod, a potent inducer of local cytokine release, has been approved for treatment of genital warts. A topical compound composed of green tea extracts (sinecatechins) is also available. Conventional and laser surgical procedures may be required for recalcitrant warts. Recurrence of warts appears to be common with all these modalities. A highly effective vaccine for selected types of HPV has been approved by the FDA, and its use is reported to reduce the incidence of anogenital and cervical carcinoma.

HERPES SIMPLEX

See Chap. 192.

HERPES ZOSTER

See Chap. 193.

ACNE

ACNE VULGARIS

Acne vulgaris is a self-limited disorder primarily of teenagers and young adults, although perhaps 10–20% of adults may continue to experience some form of the disorder. The permissive factor for the expression of the disease in adolescence is the increase in sebum production by sebaceous glands with puberty. Small cysts, called *comedones*, form in hair follicles due to blockage of the follicular orifice by retention of keratinous material and sebum. The activity of bacteria (*Cutibacterium acnes*) within the comedones releases free fatty acids from sebum, causes inflammation within the cyst, and results in rupture of the cyst wall. An inflammatory foreign-body reaction develops as result of extrusion of oily and keratinous debris from the cyst.

The clinical hallmark of acne vulgaris is the comedone, which may be closed (*whitehead*) or open (*blackhead*). Closed comedones appear as 1- to 2-mm pebbly white papules, which are accentuated when the skin is stretched. They are the precursors of inflammatory lesions of acne vulgaris. The contents of closed comedones are not easily expressed. Open comedones, which rarely result in inflammatory acne lesions, have a dilated follicular orifice and are filled with easily expressible oxidized, darkened, oily debris. Comedones are usually accompanied by inflammatory lesions: papules, pustules, or nodules.

The earliest lesions seen in adolescence are generally mildly inflamed or noninflammatory comedones on the forehead. Subsequently, more typical inflammatory lesions develop on the cheeks, nose, and chin **(Fig. 57-7)**. The most common location for acne is the face, but involvement of the chest and back is common. Most disease remains mild and does not lead to scarring. A small number of patients develop large inflammatory cysts and nodules, which may drain and result in significant scarring. Regardless of the severity, acne may affect a patient's quality of life. With adequate treatment, this effect may be transient. In the case of severe, scarring acne, the effects can be permanent and profound. Early therapeutic intervention in severe acne is essential.

FIGURE 57-7

Acne vulgaris. An example of acne vulgaris with inflammatory papules, pustules, and comedones. *(Courtesy of Kalman Watsky, MD; with permission.)*

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Acne vulgaris. An example of acne vulgaris with inflammatory papules, pustules, and comedones. *(Courtesy of Kalman Watsky, MD; with permission.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Exogenous and endogenous factors can alter the expression of acne vulgaris. Friction and trauma (from headbands or chin straps of athletic helmets), application of comedogenic topical agents (cosmetics or hair preparations), or chronic topical exposure to certain industrial compounds may elicit or aggravate acne. Glucocorticoids, topical or systemic, may also elicit acne. Other systemic medications such as progestin-only contraception, lithium, isoniazid, androgenic steroids, halogens, phenytoin, and phenobarbital may produce acneiform eruptions or aggravate preexisting acne. Genetic factors and polycystic ovary disease may also play a role.

TREATMENT

Acne Vulgaris

Treatment of acne vulgaris is directed toward elimination of comedones by normalizing follicular keratinization and decreasing sebaceous gland activity, the population of *C. acnes*, and inflammation. Minimal to moderate pauci-inflammatory disease may respond adequately to local therapy alone. Although areas affected with acne should be kept clean, overly vigorous scrubbing may aggravate acne due to mechanical rupture of comedones. Topical agents such as retinoic acid, benzoyl peroxide, or salicylic acid may alter the pattern of epidermal desquamation, preventing the formation of comedones and aiding in the resolution of preexisting cysts. Topical antibacterial agents (such as benzoyl peroxide, azelaic acid, erythromycin, clindamycin, or dapsone) are also useful adjuncts to therapy. Topical antibiotics (erythromycin and clindamycin) should be used in combination with benzoyl peroxide to prevent development of bacterial resistance.

Patients with moderate to severe acne with a prominent inflammatory component will benefit from the addition of systemic therapy, such as minocycline or doxycycline in doses of 100 mg bid or in lower dose, extended-release preparations. Such antibiotics appear to have anti-inflammatory effects independent of their antibacterial effects. Female patients who do not respond to oral antibiotics may benefit from hormonal therapy. Several oral contraceptives are now approved by the FDA for use in the treatment of acne vulgaris. Spironolactone is emerging as a safe, effective, and durable antiandrogen treatment in women.

Patients with severe nodulocystic acne unresponsive to the therapies discussed above may benefit from treatment with the synthetic retinoid isotretinoin. Dosing is weight-based and cumulative, with duration of therapy dictated by summative dose or acne lesion remission. Results are excellent in appropriately selected patients. Its use is highly regulated due to its potential for severe adverse events, primarily teratogenicity and depression. In addition, patients receiving this medication develop dry skin and cheilitis and must be followed for development of hypertriglyceridemia.

At present, prescribers must enroll in a program designed to prevent pregnancy and adverse events while patients are taking isotretinoin. These measures are imposed to ensure that all prescribers are familiar with the risks of isotretinoin, that all female patients have two negative pregnancy tests prior to initiation of therapy and a negative pregnancy test prior to each refill, and that all patients have been warned about the risks associated with isotretinoin.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

affected are Caucasians of northern European background, but rosacea also occurs in patients with dark skin. Rosacea is seen almost exclusively in adults, only rarely affecting patients <30 years old. Rosacea is more common in women, but those most severely affected are men. It is characterized by

ACNE ROSACEA

Acne rosacea, commonly referred to simply as *rosacea*, is an inflammatory disorder predominantly affecting the central face. Persons most often affected are Caucasians of northern European background, but rosacea also occurs in patients with dark skin. Rosacea is seen almost exclusively in adults, only rarely affecting patients <30 years old. Rosacea is more common in women, but those most severely affected are men. It is characterized by the presence of erythema, telangiectasias, and superficial pustules **(Fig. 57-8)** but is not associated with the presence of comedones. Rosacea rarely involves the chest or back.

FIGURE 57-8

Acne rosacea. Prominent facial erythema, telangiectasia, scattered papules, and small pustules are seen in this patient with acne rosacea. *(Courtesy of Robert Swerlick, MD; with permission.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

There is a relationship between the tendency for facial flushing and the subsequent development of acne rosacea. Often, individuals with rosacea initially demonstrate a pronounced flushing reaction. This may be in response to heat, emotional stimuli, alcohol, hot drinks, or spicy foods. As the disease progresses, the flush persists longer and longer and may eventually become permanent. Papules, pustules, and telangiectasias can become superimposed on the persistent flush. Rosacea of very long standing may lead to connective tissue overgrowth, particularly of the nose (*rhinophyma*). Rosacea may also be complicated by various inflammatory disorders of the eye, including keratitis, blepharitis, iritis, and recurrent chalazion. These ocular problems are potentially sight-threatening and warrant ophthalmologic evaluation.

TREATMENT

Acne Rosacea

Acne rosacea can be treated topically or systemically. Mild disease often responds to topical preparations of metronidazole, sodium sulfacetamide, azelaic acid, ivermectin, brimonidine, or oxymetazoline. More severe disease requires oral tetracyclines in subantimicrobial, modified-release preparations. Residual telangiectasia may respond to laser therapy. Topical glucocorticoids, especially potent agents, should be avoided because chronic use of these preparations may elicit rosacea. Application of topical agents to the skin is not effective treatment for ocular disease.

SKIN DISEASES AND SMALLPOX VACCINATION

Although smallpox vaccinations were discontinued several decades ago for the general population, they are still required for certain military personnel and first responders. In the absence of a bioterrorism attack and a real or potential exposure to smallpox, such vaccination is contraindicated in persons with a history of skin diseases, such as AD, eczema, and psoriasis, who have a higher incidence of adverse events associated with smallpox vaccination. In the case of such exposure, the risk of smallpox infection outweighs that of adverse events from the vaccine **(Chap. S3)**.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

persons with a history of skin diseases, such as AD, eczema, and psoriasis, who have a higher incidence of adverse events associated with smallpox vaccination. In the case of such exposure, the risk of smallpox infection outweighs that of adverse events from the vaccine **(Chap. S3)**.

FURTHER READING

Bolognia JL et al (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.

James WD et al (eds): *Andrew's Diseases of the Skin Clinical Dermatology*, 13th ed. Philadelphia, Elsevier, 2020.

Kang S et al (eds): *Fitzpatrick's Dermatology in General Medicine*, 9th ed. New York, McGraw-Hill, 2019.

Wolff K et al (eds): *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 8th ed. New York, McGraw-Hill, 2017.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 57: Eczema, Psoriasis, Cutaneous Infections, Acne, and Other Common Skin Disorders, Leslie P. Lawley; Justin Cheeley; Robert A. Swerlick
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 58: Skin Manifestations of Internal Disease

Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman

INTRODUCTION

It is a generally accepted concept in medicine that the skin can develop signs of internal disease. Therefore, in textbooks of medicine, one finds a chapter describing in detail the major systemic disorders that can be identified by cutaneous signs. The underlying assumption of such a chapter is that the clinician has been able to identify the specific disorder in the patient and needs only to read about it in the textbook. In reality, concise differential diagnoses and the identification of these disorders are actually difficult for the nondermatologist because he or she is not well-versed in the recognition of cutaneous lesions or their spectrum of presentations. Therefore, this chapter covers this particular topic of cutaneous medicine not by simply focusing on individual diseases, but by describing the various presenting clinical signs and symptoms that point to specific disorders. Concise differential diagnoses will be generated in which the significant diseases will be distinguished from the more common cutaneous disorders that have minimal or no significance with regard to associated internal disease. The latter disorders are reviewed in table form and always need to be excluded when considering the former. For a detailed description of individual diseases, the reader should consult a dermatologic text.

PAPULOSQUAMOUS SKIN LESIONS

(Table 58-1) When an eruption is characterized by elevated lesions, either papules (<1 cm) or plaques (>1 cm), in association with scale, it is referred to as *papulosquamous*. The most common papulosquamous diseases—*tinea*, *psoriasis*, *pityriasis rosea*, and *lichen planus*—are primary cutaneous disorders **(Chap. 57)**. When psoriatic lesions are accompanied by arthritis, the possibility of psoriatic arthritis or reactive arthritis should be considered. A history of oral ulcers, conjunctivitis, uveitis, and/or urethritis points to the latter diagnosis. Lithium, beta blockers, anti-PD-1/PD-L1 antibodies, HIV or streptococcal infections, and a rapid taper of systemic glucocorticoids are known to exacerbate psoriasis; despite being used to treat psoriasis, tumor necrosis factor (TNF) inhibitors can also induce psoriatic lesions. Comorbidities in patients with psoriasis include cardiovascular disease and metabolic syndrome.

TABLE 58-1

Selected Causes of Papulosquamous Skin Lesions

1. Primary cutaneous disorders
   a. Tinea[a]—widespread disease may be sign of immunosuppression
   b. Psoriasis[a]—widespread or resistant disease may be sign of HIV infection
   c. Pityriasis rosea[a]
   d. Lichen planus[a]
   e. Parapsoriasis, small plaque and large plaque
   f. Bowen's disease (squamous cell carcinoma in situ)[b]
2. Drugs
3. Systemic diseases
   a. Lupus erythematosus, primarily subacute or chronic (discoid) lesions[c]
   b. Cutaneous T-cell lymphoma, in particular, mycosis fungoides[d]
   c. Secondary syphilis
   d. Reactive arthritis
   e. Sarcoidosis[e]—with scale less common than without scale
   f. Bazex syndrome (acrokeratosis paraneoplastica)[f]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

[a]Discussed in detail in **Chap. 57**; cardiovascular disease and the metabolic syndrome are comorbidities in psoriasis; primarily in Europe, hepatitis C virus is associated with oral lichen planus. [b]Associated with chronic sun exposure more often than exposure to arsenic; usually one or a few lesions. [c]See also Red Lesions

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 58-1

Selected Causes of Papulosquamous Skin Lesions

1. Primary cutaneous disorders
   a. Tinea[a]—widespread disease may be sign of immunosuppression
   b. Psoriasis[a]—widespread or resistant disease may be sign of HIV infection
   c. Pityriasis rosea[a]
   d. Lichen planus[a]
   e. Parapsoriasis, small plaque and large plaque
   f. Bowen's disease (squamous cell carcinoma in situ)[b]
2. Drugs
3. Systemic diseases
   a. Lupus erythematosus, primarily subacute or chronic (discoid) lesions[c]
   b. Cutaneous T-cell lymphoma, in particular, mycosis fungoides[d]
   c. Secondary syphilis
   d. Reactive arthritis
   e. Sarcoidosis[e]—with scale less common than without scale
   f. Bazex syndrome (acrokeratosis paraneoplastica)[f]

[a]Discussed in detail in **Chap. 57**; cardiovascular disease and the metabolic syndrome are comorbidities in psoriasis; primarily in Europe, hepatitis C virus is associated with oral lichen planus. [b]Associated with chronic sun exposure more often than exposure to arsenic; usually one or a few lesions. [c]See also Red Lesions in "Papulonodular Skin Lesions." [d]Also cutaneous lesions of HTLV-1-associated adult T-cell leukemia/lymphoma. [e]See also Red-Brown Lesions in "Papulonodular Skin Lesions." [f]Psoriasiform lesions of the helices, nose, and acral sites; squamous cell carcinoma of the upper aerodigestive tract most common underlying malignancy.

Abbreviation: HIV, human immunodeficiency virus.

Whenever the clinical diagnosis of pityriasis rosea or lichen planus is made, it is important to review the patient's medications because the eruption may resolve by simply discontinuing the offending agent. Pityriasis rosea–like drug eruptions are seen most commonly with beta blockers, angiotensin-converting enzyme (ACE) inhibitors, and metronidazole, whereas the drugs that can produce a lichenoid eruption include thiazides, antimalarials, quinidine, beta blockers, TNF inhibitors, anti-PD-1/PD-L1 antibodies, and ACE inhibitors. In some populations (e.g., Europeans), there is a higher prevalence of hepatitis C viral infection in patients with oral lichen planus. Lichen planus–like lesions are also observed in chronic graft-versus-host disease.

In its early stages, the mycosis fungoides (MF) form of *cutaneous T-cell lymphoma* (CTCL) may be confused with eczema or psoriasis, but it often eventually fails to respond to appropriate therapy for those inflammatory diseases. MF can develop within lesions of large-plaque parapsoriasis and is suggested by an increase in the thickness of the lesions. The diagnosis of MF is established by skin biopsy in which collections of atypical T lymphocytes are found in the epidermis and dermis. As the disease progresses, cutaneous tumors and lymph node involvement may appear.

In *secondary syphilis*, there are scattered pink to red-brown papules with thin scale. The eruption often involves the palms and soles and can resemble pityriasis rosea. Associated findings are helpful in making the diagnosis and include nonscarring alopecia, annular plaques on the face, mucous patches, condyloma lata (broad-based and moist), and lymphadenopathy, as well as malaise, fever, headache, and myalgias. The interval between the primary chancre and the secondary stage is usually 4–8 weeks, and spontaneous resolution without appropriate therapy occurs.

ERYTHRODERMA

(Table 58-2) *Erythroderma* is the term used when the majority of the skin surface is erythematous (red in color). There may be associated scale, erosions, or pustules as well as shedding of the hair and nails. Potential systemic manifestations include fever, chills, hypothermia, reactive

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

red scaly plaques on the elbows and knees would point to psoriasis. It is also important to examine the skin carefully for a migration of the erythema and associated secondary changes such as pustules or erosions. Migratory waves of erythema studded with superficial pustules are seen in *pustular*

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

(Table 58-2) *Erythroderma* is the term used when the majority of the skin surface is erythematous (red in color). There may be associated scale, erosions, or pustules as well as shedding of the hair and nails. Potential systemic manifestations include fever, chills, hypothermia, reactive lymphadenopathy, peripheral edema, hypoalbuminemia, and high-output cardiac failure. The major etiologies of erythroderma are (1) cutaneous diseases such as psoriasis and dermatitis (Table 58-3); (2) drugs; (3) systemic diseases, most commonly CTCL; and (4) idiopathic. In the first three groups, the location and description of the initial lesions, prior to the development of the erythroderma, aid in the diagnosis. For example, a history of red scaly plaques on the elbows and knees would point to psoriasis. It is also important to examine the skin carefully for a migration of the erythema and associated secondary changes such as pustules or erosions. Migratory waves of erythema studded with superficial pustules are seen in *pustular psoriasis*.

TABLE 58-2

Causes of Erythroderma

1. Primary cutaneous disorders
   a. Psoriasis[a]
   b. Dermatitis (atopic > contact >> stasis [with autosensitization] or seborrheic [primarily infants])[a]
   c. Pityriasis rubra pilaris
2. Drugs
3. Systemic diseases
   a. Cutaneous T-cell lymphoma (Sézary syndrome, erythrodermic mycosis fungoides)
   b. Other lymphomas
   c. Rarely, late-stage solid tumors
4. Idiopathic (usually older men)

[a]Discussed in detail in Chap. 57.

TABLE 58-3

Erythroderma (Primary Cutaneous Disorders)

	INITIAL LESIONS	LOCATION OF INITIAL LESIONS	OTHER FINDINGS	DIAGNOSTIC AIDS	TREATMENT
Psoriasis[a]	Pink-red, silvery scale, sharply demarcated	Elbows, knees, scalp, presacral area, intergluteal fold	Nail dystrophy (e.g., pits, oil drop sign), arthritis, pustules, SAPHO syndrome[b]	Skin biopsy	Topical glucocorticoids, vitamin D analogs; UV-B (narrowband) > PUVA; oral retinoids; MTX; anti-TNF agents, anti-IL-12/23 Ab, anti-IL-23 Ab, anti-IL-17A or -IL-17 receptor A Ab; apremilast; cyclosporine
Dermatitis[a]					
Atopic	Acute: Erythema, fine scale, crust, indistinct borders, excoriations Chronic:	Antecubital and popliteal fossae, neck, hands, eyelids	Pruritus Personal and/or family history of atopy, including asthma, allergic rhinitis or conjunctivitis, and atopic dermatitis Exclude secondary infection with *Staphylococcus aureus* or HSV	Skin biopsy	Topical glucocorticoids, tacrolimus, pimecrolimus, tar, crisaborole, and antipruritics; oral antihistamines for sedation; open wet dressings; UV-B ± UV-A > PUVA; anti-IL-4/13 Ab; oral/IM glucocorticoids (short-term); MTX; mycophenolate mofetil; azathioprine;
	markings), excoriations				

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

	Chronic: Lichenification (increased skin markings), excoriations		*Staphylococcus aureus* or HSV Exclude superimposed irritant or allergic contact dermatitis		mycophenolate mofetil; azathioprine; cyclosporine Topical or oral antibiotics
Contact	Local: Erythema, crusting, vesicles, and bullae Systemic: Erythema, fine scale, crust	Depends on offending agent Generalized vs major intertriginous zones (especially groin)	Irritant—onset often within hours Allergic—delayed-type hypersensitivity; lag time of 48 h with rechallenge Patient has history of allergic contact dermatitis to topical agent and then receives systemic medication that is structurally related, e.g., formaldehyde (skin), aspartame (oral)	Patch testing; repeat open application test Patch testing	Remove irritant or allergen; topical glucocorticoids; oral antihistamines; oral/IM glucocorticoids (short-term) Same as local
Seborrheic (rare in adults)	Pink-red to pink-orange, greasy scale	Scalp, nasolabial folds, eyebrows, intertriginous zones	Flares with stress, HIV infection Associated with Parkinson's disease	Skin biopsy	Topical glucocorticoids and imidazoles
Stasis (with autosensitization)	Erythema, crusting, excoriations	Lower extremities	Pruritus, lower extremity edema, varicosities, hemosiderin deposits, lipodermatosclerosis History of venous ulcers, thrombophlebitis, and/or cellulitis Exclude cellulitis Exclude superimposed contact dermatitis, e.g., topical neomycin	Skin biopsy	Topical glucocorticoids; open wet dressings; leg elevation; pressure stockings; pressure wraps if associated ulcers
Pityriasis rubra pilaris	Orange-red (salmon-colored), perifollicular papules	Generalized, but characteristic "skip" areas of normal skin	Wax-like palmoplantar keratoderma Exclude cutaneous T-cell lymphoma	Skin biopsy	Isotretinoin or acitretin; MTX; anti-IL-12/23 Ab, anti-IL-23 Ab, anti-TNF agents, anti-IL-17A or -IL-17 receptor A Ab

[a]Discussed in detail in **Chap. 57**. [b]SAPHO syndrome occurs more commonly in patients with palmoplantar pustulosis than in those with erythrodermic psoriasis.

Abbreviations: Ab, antibody; HSV, herpes simplex virus; IL, interleukin; IM, intramuscular; MTX, methotrexate; PUVA, psoralens plus ultraviolet A irradiation; SAPHO, *s*ynovitis, *a*cne, *p*ustulosis, *h*yperostosis, and *o*steitis (a subtype is chronic recurrent multifocal osteomyelitis); TNF, tumor necrosis factor; UV-A, ultraviolet A irradiation; UV-B, ultraviolet B irradiation.

Drug-induced erythroderma may begin as an exanthematous (morbilliform) eruption **(Chap. 60)** or may arise as diffuse erythema. A number of drugs

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

hypersensitivity syndrome (DIHS). In addition, these reactions, especially to aromatic anticonvulsants, can lead to a pseudolymphoma syndrome with adenopathy and circulating atypical lymphocytes, while reactions to allopurinol may be accompanied by gastrointestinal bleeding

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Drug-induced erythroderma may begin as an exanthematous (morbilliform) eruption **(Chap. 60)** or may arise as diffuse erythema. A number of drugs can produce an erythroderma, including penicillins, sulfonamides, aromatic anticonvulsants (e.g., carbamazepine, phenytoin), and allopurinol. Fever and peripheral eosinophilia often accompany the eruption, and there may also be facial swelling, hepatitis, myocarditis, thyroiditis, and allergic interstitial nephritis; this constellation is frequently referred to as *drug reaction with eosinophilia and systemic symptoms* (DRESS) or *drug-induced hypersensitivity syndrome* (DIHS). In addition, these reactions, especially to aromatic anticonvulsants, can lead to a pseudolymphoma syndrome with adenopathy and circulating atypical lymphocytes, while reactions to allopurinol may be accompanied by gastrointestinal bleeding.

The most common malignancy that is associated with erythroderma is CTCL; in some series, up to 25% of the cases of erythroderma were due to CTCL. The patient may progress from isolated plaques and tumors, but more commonly, the erythroderma is present throughout the course of the disease (Sézary syndrome). In Sézary syndrome, there are circulating clonal atypical T lymphocytes, pruritus, and lymphadenopathy. In cases of erythroderma where there is no apparent cause (idiopathic), longitudinal evaluation is mandatory to monitor for the possible development of CTCL.

ALOPECIA

(Table 58-4) The two major forms of alopecia are scarring and nonscarring. *Scarring alopecia* is associated with fibrosis, inflammation, and loss of hair follicles. A smooth scalp with a decreased number of follicular openings is usually observed clinically, but in some patients, the changes are seen only in biopsy specimens from affected areas. In *nonscarring alopecia*, the hair shafts are absent or miniaturized, but the hair follicles are preserved, explaining the reversible nature of nonscarring alopecia.

TABLE 58-4

Causes of Alopecia

I. Nonscarring alopecia
- A. Primary cutaneous disorders
  1. Androgenetic alopecia (female pattern, male pattern)
  2. Telogen effluvium
  3. Alopecia areata
  4. Tinea capitis
  5. Traumatic alopecia[a]
  6. Psoriasiform alopecia, including TNF inhibitor–induced
- B. Drugs
  1. Telogen effluvium—see text for most common causes
  2. Anagen effluvium—chemotherapeutic agents (e.g., anthracyclines)
- C. Systemic diseases
  1. Systemic lupus erythematosus
  2. Secondary syphilis
  3. Hypothyroidism
  4. Hyperthyroidism
  5. Hypopituitarism
  6. Deficiencies of protein, biotin, zinc, and perhaps iron

II. Scarring alopecia
- A. Primary cutaneous disorders
  1. Cutaneous lupus (chronic discoid lesions)[b]
  2. Lichen planus, including frontal fibrosing alopecia
  3. Central centrifugal cicatricial alopecia
  4. Folliculitis decalvans
  5. Dissecting cellulitis
  6. Linear morphea (linear scleroderma)[c]
- B. Drugs
  1. Chemotherapeutic agents (e.g., taxanes, busulfan)
  2. Sarcoidosis
  3. Cutaneous metastases

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

(Table 58-4) The two major forms of alopecia are scarring and nonscarring. *Scarring alopecia* is associated with fibrosis, inflammation, and loss of

TABLE 58-4

Causes of Alopecia

I. Nonscarring alopecia
  A. Primary cutaneous disorders
    1. Androgenetic alopecia (female pattern, male pattern)
    2. Telogen effluvium
    3. Alopecia areata
    4. Tinea capitis
    5. Traumatic alopecia[a]
    6. Psoriasiform alopecia, including TNF inhibitor–induced
  B. Drugs
    1. Telogen effluvium—see text for most common causes
    2. Anagen effluvium—chemotherapeutic agents (e.g., anthracyclines)
  C. Systemic diseases
    1. Systemic lupus erythematosus
    2. Secondary syphilis
    3. Hypothyroidism
    4. Hyperthyroidism
    5. Hypopituitarism
    6. Deficiencies of protein, biotin, zinc, and perhaps iron
II. Scarring alopecia
  A. Primary cutaneous disorders
    1. Cutaneous lupus (chronic discoid lesions)[b]
    2. Lichen planus, including frontal fibrosing alopecia
    3. Central centrifugal cicatricial alopecia
    4. Folliculitis decalvans
    5. Dissecting cellulitis
    6. Linear morphea (linear scleroderma)[c]
  B. Drugs
    1. Chemotherapeutic agents (e.g., taxanes, busulfan)
  C. Systemic diseases
    1. Discoid lesions in the setting of systemic lupus erythematosus[b]
    2. Sarcoidosis
    3. Cutaneous metastases

[a]Most patients with trichotillomania or early stages of traction alopecia and some patients with pressure-induced alopecia. [b]While the majority of patients with discoid lesions have only cutaneous disease, these lesions do represent one of the criteria in the European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) [2019] and ACR [1982] classification schemes for systemic lupus erythematosus. [c]Can involve underlying muscles and osseous structures, and rarely in linear morphea of the frontal scalp (*en coup de sabre*), there is involvement of the meninges and brain.

The most common causes of nonscarring alopecia include *androgenetic alopecia*, *telogen effluvium*, *alopecia areata*, *tinea capitis*, and the early phase of *traumatic alopecia* (Table 58-5). In women with androgenetic alopecia, an elevation in circulating levels of androgens may be seen as a result of ovarian or adrenal gland dysfunction or neoplasm. When there are signs of virilization, such as a deepened voice and/or enlarged clitoris, the possibility of an ovarian or adrenal gland tumor should be considered.

TABLE 58-5

Nonscarring Alopecia (Primary Cutaneous Disorders)

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Telogen	Diffuse shedding of normal hairs	Stress causes more of the asynchronous growth	Observation; discontinue any drugs that have

TABLE 58-5

Nonscarring Alopecia (Primary Cutaneous Disorders)

	CLINICAL CHARACTERISTICS	PATHOGENESIS	TREATMENT
Telogen effluvium	Diffuse shedding of normal hairs Follows major stress (high fever, severe infection) or change in hormone levels (postpartum) Reversible without treatment	Stress causes more of the asynchronous growth cycles of individual hairs to become synchronous; therefore, larger numbers of growing (anagen) hairs simultaneously enter the dying (telogen) phase	Observation; discontinue any drugs that have alopecia as a side effect; must exclude underlying metabolic causes, e.g., hypothyroidism, hyperthyroidism
Androgenetic alopecia (male pattern; female pattern)	Miniaturization of hairs along the midline of the scalp Recession of the anterior scalp line in men and some women	Increased sensitivity of affected hairs to the effects of androgens—most common Increased levels of circulating androgens (ovarian or adrenal source in women)—less common	If no evidence of hyperandrogenemia, then topical minoxidil; finasteride[a]; spironolactone (women); hair transplant; low-dose oral minoxidil
Alopecia areata	Well-circumscribed, circular areas of hair loss, 2–5 cm in diameter In extensive cases, coalescence of lesions and/or involvement of other hair-bearing surfaces of the body Pitting or sandpapered appearance of the nails	The germinative zones of the hair follicles are surrounded by T lymphocytes Occasional associated diseases: hyperthyroidism, hypothyroidism, vitiligo, Down syndrome	Topical anthralin or tazarotene; intralesional glucocorticoids; topical contact sensitizers; JAK inhibitors
Tinea capitis	Varies from scaling with minimal hair loss to discrete patches with "black dots" (sites of broken infected hairs) to boggy plaque with pustules (kerion)[b]	Invasion of hairs by dermatophytes, most commonly *Trichophyton tonsurans*	Oral griseofulvin or terbinafine plus 2.5% selenium sulfide or ketoconazole shampoo; examine family members
Traumatic alopecia[c]	Broken hairs, often of varying lengths Irregular outline in trichotillomania and traction alopecia Fringe sign in traction alopecia	Traction with curlers, rubber bands, tight braiding Exposure to heat or chemicals (e.g., hair straighteners) Mechanical pulling (trichotillomania)	Discontinuation of offending hair style or chemical treatments; diagnosis of trichotillomania may require observation of shaved hairs (for growth) or biopsy, possibly followed by psychotherapy

[a]To date, Food and Drug Administration–approved for men. [b]Scarring alopecia can occur at sites of kerions. [c]May also be scarring, especially late-stage traction alopecia.

Exposure to various drugs can also cause diffuse hair loss, usually by inducing a telogen effluvium. An exception is the anagen effluvium observed with chemotherapeutic agents such as daunorubicin. Alopecia is a side effect of the following drugs: warfarin, heparin, propylthiouracil, carbimazole, isotretinoin, acitretin, lithium, beta blockers, interferons, colchicine, and amphetamines. Fortunately, spontaneous regrowth usually follows discontinuation of the offending agent.

Less commonly, nonscarring alopecia is associated with *lupus erythematosus* and *secondary syphilis.* In systemic lupus, there are two forms of alopecia—one is scarring secondary to discoid lesions (see below), and the other is nonscarring. The latter form coincides with flares of systemic disease and may involve the entire scalp or just the frontal scalp, with the appearance of multiple short hairs ("lupus hairs") as a sign of initial

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Scarring alopecia is more frequently the result of a primary cutaneous disorder such as *lichen planus*, *chronic cutaneous (discoid) lupus*, *central centrifugal cicatricial alopecia*, *folliculitis decalvans*, or *linear scleroderma (morphea)* than it is a sign of systemic disease. Although the scarring lesions

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

disease and may involve the entire scalp or just the frontal scalp, with the appearance of multiple short hairs ("lupus hairs") as a sign of initial regrowth. Scattered, poorly circumscribed patches of alopecia with a "moth-eaten" appearance are a manifestation of the secondary stage of syphilis. Diffuse thinning of the hair is also associated with hypothyroidism and hyperthyroidism (Table 58-4).

Scarring alopecia is more frequently the result of a primary cutaneous disorder such as *lichen planus*, *chronic cutaneous (discoid) lupus*, *central centrifugal cicatricial alopecia*, *folliculitis decalvans*, or *linear scleroderma* (*morphea*) than it is a sign of systemic disease. Although the scarring lesions of *discoid lupus* can be seen in patients with systemic lupus, in the majority of patients, the disease process is limited to the skin. Less common causes of scarring alopecia include *sarcoidosis* (see "Papulonodular Skin Lesions," below), chemotherapeutic agents, and *cutaneous metastases*.

In the early phases of discoid lupus, lichen planus, and folliculitis decalvans, there are circumscribed areas of alopecia. Fibrosis and subsequent loss of hair follicles are observed primarily in the center of these alopecic patches, whereas the inflammatory process is most prominent at the periphery. The areas of active inflammation in discoid lupus are erythematous with scale, whereas the areas of previous inflammation are often hypopigmented with a rim of hyperpigmentation. In lichen planus, perifollicular macules at the periphery are usually violet-colored. A complete examination of the skin and oral mucosa combined with a biopsy and direct immunofluorescence microscopy of inflamed skin will aid in distinguishing these two entities. The peripheral active lesions in folliculitis decalvans are follicular pustules; these patients can develop a reactive arthritis.

FIGURATE SKIN LESIONS

(Table 58-6) In *figurate eruptions*, the lesions form rings and arcs that are usually erythematous but can be skin-colored to brown. Most commonly, they are due to primary cutaneous diseases such as *tinea*, *urticaria*, *granuloma annulare*, and *erythema annulare centrifugum* **(Chaps. 57 and 59)**. An underlying systemic illness is found in a second, less common group of migratory annular erythemas. It includes *erythema migrans*, *erythema gyratum repens*, *erythema marginatum*, and *necrolytic migratory erythema*.

TABLE 58-6

Causes of Figurate Skin Lesions

I. Primary cutaneous disorders
   - A. Tinea
   - B. Urticaria (primary in ≥90% of patients)
   - C. Granuloma annulare
   - D. Erythema annulare centrifugum
   - E. Psoriasis, annular pustular psoriasis
   - F. Interstitial granulomatous drug reaction

II. Systemic diseases
   - A. Migratory
     1. Erythema migrans (CDC case definition is ≥5 cm in diameter)
     2. Urticaria (≤10% of patients)
     3. Erythema gyratum repens
     4. Erythema marginatum
     5. Pustular psoriasis (generalized and annular forms)
     6. Necrolytic migratory erythema (glucagonoma syndrome)[a]
   - B. Nonmigratory (may slowly expand)
     1. Subacute cutaneous LE, LE tumidus
     2. Sarcoidosis
     3. Leprosy (borderline, tuberculoid)
     4. Secondary syphilis (especially the face)
     5. Cutaneous T-cell lymphoma (especially mycosis fungoides)
     6. Interstitial granulomatous dermatitis[b]
     7. Annular erythema of Sjögren's syndrome

Abbreviations: CDC, Centers for Disease Control and Prevention; LE, lupus erythematosus

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 58-6

Causes of Figurate Skin Lesions

I. Primary cutaneous disorders
   A. Tinea
   B. Urticaria (primary in ≥90% of patients)
   C. Granuloma annulare
   D. Erythema annulare centrifugum
   E. Psoriasis, annular pustular psoriasis
   F. Interstitial granulomatous drug reaction
II. Systemic diseases
   A. Migratory
      1. Erythema migrans (CDC case definition is ≥5 cm in diameter)
      2. Urticaria (≤10% of patients)
      3. Erythema gyratum repens
      4. Erythema marginatum
      5. Pustular psoriasis (generalized and annular forms)
      6. Necrolytic migratory erythema (glucagonoma syndrome)[a]
   B. Nonmigratory (may slowly expand)
      1. Subacute cutaneous LE, LE tumidus
      2. Sarcoidosis
      3. Leprosy (borderline, tuberculoid)
      4. Secondary syphilis (especially the face)
      5. Cutaneous T-cell lymphoma (especially mycosis fungoides)
      6. Interstitial granulomatous dermatitis[b]
      7. Annular erythema of Sjögren's syndrome

[a]Migratory erythema with erosions; favors lower extremities and girdle area. [b]Underlying diseases include rheumatoid arthritis, LE, and granulomatosis with polyangiitis.

Abbreviations: CDC, Centers for Disease Control and Prevention; LE, lupus erythematosus.

In erythema gyratum repens, one sees numerous mobile concentric arcs and wavefronts that resemble the grain in wood. A search for an underlying malignancy is mandatory in a patient with this eruption. Erythema migrans is the cutaneous manifestation of Lyme disease, which is caused by the spirochete *Borrelia burgdorferi*. In the initial stage (3–30 days after tick bite), a single annular lesion is usually seen, which can expand to ≥10 cm in diameter. Within several days, up to half of the patients develop multiple smaller erythematous lesions at sites distant from the bite. Associated symptoms include fever, headache, photophobia, myalgias, arthralgias, and malar rash. Erythema marginatum is seen in patients with rheumatic fever, primarily on the trunk. Lesions are pink-red in color, flat to minimally elevated, and transient.

There are additional cutaneous diseases that present as annular eruptions but lack an obvious migratory component. Examples include *CTCL*, *subacute cutaneous lupus*, *secondary syphilis*, and *sarcoidosis* (see "Papulonodular Skin Lesions," below).

ACNE

(Table 58-7) In addition to *acne vulgaris* and *acne rosacea*, the two major forms of acne **(Chap. 57)**, there are drugs and systemic diseases that can lead to acneiform eruptions.

TABLE 58-7

Causes of Acneiform Eruptions

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

A. Acne vulgaris
B. Acne rosacea

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 58-7

Causes of Acneiform Eruptions

I. Primary cutaneous disorders
  A. Acne vulgaris
  B. Acne rosacea
II. Drugs, e.g., anabolic steroids, glucocorticoids, lithium, EGFR inhibitors, HER2 inhibitors, MEK inhibitors, iodides
III. Systemic diseases
  A. Increased androgen production
    1. Adrenal origin, e.g., Cushing's disease, 21-hydroxylase deficiency
    2. Ovarian origin, e.g., polycystic ovary syndrome, ovarian hyperthecosis
  B. Cryptococcosis, disseminated
  C. Dimorphic fungal infections
  D. Behçet's disease

Abbreviation: EGFR, epidermal growth factor receptor; HER2, human epidermal growth factor receptor 2; MEK, MAP (mitogen activated protein) kinase.

Patients with the *carcinoid syndrome* have episodes of flushing of the head, neck, and sometimes the trunk. Resultant skin changes of the face, in particular telangiectasias, may mimic the clinical appearance of erythematotelangiectatic acne rosacea.

PUSTULAR LESIONS

Acneiform eruptions (see "Acne," above) and *folliculitis* represent the most common pustular dermatoses. An important consideration in the evaluation of follicular pustules is a determination of the associated pathogen, for example, normal flora (culture-negative), *Staphylococcus aureus*, *Pseudomonas aeruginosa* ("hot tub" folliculitis), *Malassezia*, dermatophytes (Majocchi's granuloma), and *Demodex* spp. Noninfectious forms of folliculitis include HIV- or immunosuppression-associated eosinophilic folliculitis and folliculitis secondary to drugs such as glucocorticoids, lithium, and epidermal growth factor receptor (EGFR) or MEK inhibitors. Administration of high-dose systemic glucocorticoids can result in a widespread eruption of follicular pustules on the trunk, characterized by lesions in the same stage of development. With regard to underlying systemic diseases, nonfollicular-based pustules are a characteristic component of pustular psoriasis (sterile) and can be seen in septic emboli of bacterial or fungal origin (see "Purpura," below). In patients with acute generalized exanthematous pustulosis (AGEP) due primarily to medications (e.g., cephalosporins), there are large areas of erythema studded with multiple sterile pustules in addition to neutrophilia.

TELANGIECTASIAS

(Table 58-8) To distinguish the various types of telangiectasias, it is important to examine the shape and configuration of the dilated blood vessels. *Linear telangiectasias* are seen on the face of patients with *actinically damaged skin* and *acne rosacea,* and they are found on the legs of patients with *venous hypertension* and first appear on the legs in *generalized essential telangiectasia.* Patients with an unusual form of *mastocytosis* (telangiectasia macularis eruptiva perstans) and the *carcinoid syndrome* (see "Acne," above) also have linear telangiectasias. Lastly, linear telangiectasias are found in areas of cutaneous inflammation. For example, longstanding lesions of discoid lupus frequently have telangiectasias within them.

TABLE 58-8

Causes of Telangiectasias

I. Primary cutaneous disorders
  A. Linear/branching
    1. Acne rosacea (face)
    2. Actinically damaged skin (face, neck, V of chest)
    3. Venous hypertension (legs)

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

  B. Poikiloderma
    1. Ionizing radiation[a]

TABLE 58-8

Causes of Telangiectasias

I. Primary cutaneous disorders
  A. Linear/branching
    1. Acne rosacea (face)
    2. Actinically damaged skin (face, neck, V of chest)
    3. Venous hypertension (legs)
    4. Generalized essential telangiectasia
    5. Cutaneous collagenous vasculopathy
    6. Within basal cell carcinomas or cutaneous lymphoma
  B. Poikiloderma
    1. Ionizing radiation[a]
  C. Spider angioma
    1. Idiopathic
    2. Pregnancy
II. Systemic diseases
  A. Linear/branching
    1. Carcinoid (head, neck, upper trunk)
    2. Ataxia-telangiectasia (bulbar conjunctivae, head and neck)
    3. Mastocytosis (within lesions)
  B. Poikiloderma
    1. Dermatomyositis, lupus erythematosus
    2. Mycosis fungoides, patch stage
    3. Genodermatoses, e.g., xeroderma pigmentosum, Kindler syndrome
  C. Mat
    1. Systemic sclerosis (scleroderma)
  D. Nailfold
    1. Lupus erythematosus
    2. Systemic sclerosis (scleroderma)
    3. Dermatomyositis
    4. Hereditary hemorrhagic telangiectasia
  E. Papular
    1. Hereditary hemorrhagic telangiectasia
  F. Spider angioma
    1. Cirrhosis[b]

[a]Becoming less common. [b]Due to hyperestrogenic state.

Poikiloderma is a term used to describe a patch of skin with: (1) reticulated hypo- and hyperpigmentation, (2) wrinkling secondary to epidermal atrophy, and (3) telangiectasias. Poikiloderma does not imply a single disease entity—although it is becoming less common, it is seen in skin damaged by *ionizing radiation* as well as in patients with autoimmune connective tissue diseases, primarily *dermatomyositis* (DM), and rare genodermatoses (e.g., Kindler syndrome).

In *systemic sclerosis (scleroderma)*, the dilated blood vessels have a unique configuration and are known as *mat telangiectasias.* The lesions are broad macules that usually measure 2–7 mm in diameter but occasionally are larger. Mats have a polygonal or oval shape, and their erythematous color may appear uniform, but, upon closer inspection, the erythema is the result of delicate telangiectasias. The most common locations for mat telangiectasias are the face, oral mucosa, and hands—peripheral sites that are prone to intermittent ischemia. The limited form of systemic sclerosis, also referred to

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Nailfold telangiectasias are pathognomonic signs of the three major autoimmune connective tissue diseases: *lupus erythematosus*, *systemic sclerosis*, and DM. They are easily visualized by the naked eye and occur in at least two-thirds of these patients. In both DM and lupus, there is associated nailfold

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

(e.g., Kindler syndrome).

appear uniform; but, upon closer inspection, the erythema is the result of delicate telangiectasias. The most common locations for mat telangiectasias are the face, oral mucosa, and hands—peripheral sites that are prone to intermittent ischemia. The limited form of systemic sclerosis, also referred to as the CREST (*c*alcinosis cutis, *R*aynaud's phenomenon, *e*sophageal dysmotility, *s*clerodactyly, and *t*elangiectasia) variant **(Chap. 360)**, is associated with a chronic course and anticentromere antibodies. Mat telangiectasias are an important clue to the diagnosis of this variant as well as the diffuse form of systemic sclerosis because they may be the only cutaneous finding.

Nailfold telangiectasias are pathognomonic signs of the three major autoimmune connective tissue diseases: *lupus erythematosus*, *systemic sclerosis*, and *DM*. They are easily visualized by the naked eye and occur in at least two-thirds of these patients. In both DM and lupus, there is associated nailfold erythema, and in DM, the erythema is often accompanied by "ragged" cuticles and fingertip tenderness. Under 10× magnification or by dermoscopy, the blood vessels in the nailfolds of lupus patients are tortuous and resemble "glomeruli," whereas in systemic sclerosis and DM, there is a loss of capillary loops and those that remain are markedly dilated.

In *hereditary hemorrhagic telangiectasia* (Osler-Rendu-Weber disease), the lesions usually appear during adolescence (mucosal) and adulthood (cutaneous) and are most commonly seen on the mucous membranes (nasal, orolabial), face, and distal extremities, including under the nails. They represent arteriovenous (AV) malformations of the dermal microvasculature, are dark red in color, and are usually slightly elevated. When the skin is stretched over an individual lesion, an eccentric punctum with radiating legs is seen. Although the degree of systemic involvement varies in this autosomal dominant disease (due primarily to mutations in either the endoglin or activin receptor–like kinase gene), the major symptoms are recurrent epistaxis and gastrointestinal bleeding. The fact that these mucosal telangiectasias are actually AV communications helps to explain their tendency to bleed.

HYPOPIGMENTATION

(Table 58-9) Disorders of hypopigmentation are often classified as either diffuse or localized. The classic example of *diffuse hypopigmentation* is *oculocutaneous albinism* (OCA). The most common forms are due to mutations in the tyrosinase gene (type I) or the *P* gene (type II); patients with type IA OCA have a total lack of enzyme activity. At birth, different forms of OCA can appear similar—white hair, gray-blue eyes, and pink-white skin. However, the patients with no tyrosinase activity maintain this phenotype, whereas those with decreased activity will acquire some pigmentation of the eyes, hair, and skin as they age. The degree of pigment formation is also a function of racial background, and the pigmentary dilution is more readily apparent when patients are compared to their first-degree relatives. The ocular findings in OCA correlate with the degree of hypopigmentation and include decreased visual acuity, nystagmus, photophobia, strabismus, and a lack of normal binocular vision.

TABLE 58-9

Causes of Hypopigmentation

I. Primary cutaneous disorders
  A. Diffuse
    1. Generalized vitiligo[a]
  B. Localized
    1. Postinflammatory
    2. Idiopathic guttate hypomelanosis
    3. Pityriasis (tinea) versicolor
    4. Vitiligo[a]
    5. Chemical- or drug-induced leukoderma, e.g., topical imiquimod, oral imatinib
    6. Nevus depigmentosus and pigmentary mosaicism
    7. Progressive macular hypomelanosis
    8. Piebaldism[a]
II. Systemic diseases
  A. Diffuse
    1. Oculocutaneous albinism[b]
    2. Hermansky-Pudlak syndrome[b,c]
    3. Chédiak-Higashi syndrome[b,d]
    4. Phenylketonuria
    2. Melanoma-associated vitiligo-like leukoderma, immunotherapy-induced or spontaneous[e]
    3. Sarcoidosis

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 58-9

Causes of Hypopigmentation

I. Primary cutaneous disorders
  A. Diffuse
    1. Generalized vitiligo[a]
  B. Localized
    1. Postinflammatory
    2. Idiopathic guttate hypomelanosis
    3. Pityriasis (tinea) versicolor
    4. Vitiligo[a]
    5. Chemical- or drug-induced leukoderma, e.g., topical imiquimod, oral imatinib
    6. Nevus depigmentosus and pigmentary mosaicism
    7. Progressive macular hypomelanosis
    8. Piebaldism[a]
II. Systemic diseases
  A. Diffuse
    1. Oculocutaneous albinism[b]
    2. Hermansky-Pudlak syndrome[b,c]
    3. Chédiak-Higashi syndrome[b,d]
    4. Phenylketonuria
  B. Localized
    1. Systemic sclerosis (scleroderma)[e]
    2. Melanoma-associated vitiligo-like leukoderma, immunotherapy-induced or spontaneous[e]
    3. Sarcoidosis
    4. Cutaneous T-cell lymphoma (especially mycosis fungoides)
    5. Tuberculoid and indeterminate leprosy
    6. Onchocerciasis[e]
    7. Linear nevoid hypopigmentation (pigmentary mosaicism)[b,f]
    8. Incontinentia pigmenti (stage IV)
    9. Tuberous sclerosis
    10. Waardenburg syndrome and Shah-Waardenburg syndrome
    11. Vogt-Koyanagi-Harada syndrome[e]

[a]Absence of melanocytes in areas of leukoderma; congenital in piebaldism. [b]Normal number of melanocytes. [c]Platelet storage defect and restrictive lung disease secondary to deposits of ceroid-like material or immunodeficiency; due to mutations in β or δ subunit of adaptor-related protein complex 3 as well as subunits of *b*iogenesis of *l*ysosome-related *o*rganelles *c*omplex (BLOC)-1, -2, and -3. [d]Giant lysosomal granules and recurrent infections. [e]Can resemble vitiligo due to acquired complete loss of pigment. [f]Minority of patients in a nonreferral setting have systemic abnormalities (musculoskeletal, central nervous system, ocular), previously referred to as hypomelanosis of Ito.

The differential diagnosis of *localized hypomelanosis* includes the following primary cutaneous disorders: *postinflammatory hypopigmentation*, *idiopathic guttate hypomelanosis*, *pityriasis* (*tinea*) *versicolor*, *vitiligo*, *chemical- or drug-induced leukoderma*, *nevus depigmentosus* (see below), *progressive macular hypomelanosis*, and *piebaldism* **(Table 58-10)**. In this group of diseases, the areas of involvement are macules or patches with a decrease or absence of pigmentation. Patients with vitiligo also have an increased incidence of several autoimmune disorders, including Hashimoto's thyroiditis, Graves' disease, pernicious anemia, Addison's disease, uveitis, alopecia areata, chronic mucocutaneous candidiasis, and the autoimmune polyendocrine syndromes (types I and II). Diseases of the thyroid gland are the most frequently associated disorders, occurring in up to 30% of

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 58-10

Hypopigmentation (Primary Cutaneous Disorders, Localized)

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

thyroiditis, Graves' disease, pernicious anemia, Addison's disease, uveitis, alopecia areata, chronic mucocutaneous candidiasis, and the autoimmune polyendocrine syndromes (types I and II). Diseases of the thyroid gland are the most frequently associated disorders, occurring in up to 30% of patients with vitiligo. Circulating autoantibodies are often found, and the most common ones are antithyroglobulin, antimicrosomal, and antithyroid-stimulating hormone receptor antibodies.

TABLE 58-10

Hypopigmentation (Primary Cutaneous Disorders, Localized)

	CLINICAL CHARACTERISTICS	WOOD'S LAMP EXAMINATION (UV-A; PEAK = 365 NM)	SKIN BIOPSY SPECIMEN	PATHOGENESIS	TREATMENT
Postinflammatory hypopigmentation	Can develop within active lesions, as in subacute cutaneous lupus, or after the lesion fades, as in atopic dermatitis	Depends on particular disease Usually less enhancement than in vitiligo	Type of inflammatory infiltrate depends on specific disease	Block in transfer of melanin from melanocytes to keratinocytes could be secondary to edema or decrease in contact time Destruction of melanocytes if inflammatory cells attack basal layer of epidermis	Treat underlying inflammatory disease
Idiopathic guttate hypomelanosis	Common; acquired; usually 2–4 mm in diameter Shins and extensor forearms	Less enhancement than vitiligo	Abrupt decrease in epidermal melanin content	Possible somatic mutations as a reflection of aging or UV exposure	None
Pityriasis (tinea) versicolor[a]	Common disorder Upper trunk and neck (shawl-like distribution), groin Young adults Macules have fine white scale when scratched	Golden fluorescence	Hyphal forms and budding yeast in stratum corneum	Invasion of stratum corneum by the yeast *Malassezia* Yeast is lipophilic and produces C_9 and C_{11} dicarboxylic acids, which in vitro inhibit tyrosinase	Selenium sulfide 2.5% shampoo; topical imidazoles; oral triazoles
Vitiligo	Acquired; progressive Symmetric areas of complete pigment loss Periorificial—around mouth, nose, eyes, nipples, umbilicus, anus Other areas—flexor wrists, extensor distal extremities Segmental form is less common—unilateral, dermatomal-like	More apparent Chalk-white	Absence of melanocytes in well-developed lesions Mild inflammation	Autoimmune phenomenon that results in destruction of melanocytes—primarily cellular (circulating skin-homing autoreactive T cells)	Topical glucocorticoids; topical calcineurin inhibitors; UV-B (narrowband); PUVA; JAK inhibitors; transplants, if stable; depigmentation (topical MBEH), if widespread and treatment-resistant
	when associated with chemical exposure		melanocytes	phenols and catechols (germicides; rubber products)	undergo repigmentation when medication is discontinued

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

Chemical- or drug-induced leukoderma	Similar appearance to vitiligo Often begins on hands when associated with chemical exposure Satellite lesions in areas not exposed to chemicals	More apparent Chalk-white	Decreased number or absence of melanocytes	Exposure to chemicals that selectively destroy melanocytes, in particular phenols and catechols (germicides; rubber products) or ingestion of drugs such as imatinib Release of cellular antigens and activation of circulating lymphocytes may explain satellite phenomenon Possible inhibition of KIT receptor	Avoid exposure to offending agent, then treat as vitiligo Drug-induced variant may undergo repigmentation when medication is discontinued
Piebaldism	Autosomal dominant Congenital, stable White forelock Areas of amelanosis contain normally pigmented and hyperpigmented macules of various sizes Symmetric involvement of central forehead, ventral trunk, and mid regions of upper and lower extremities	Enhancement of leukoderma and hyperpigmented macules	Amelanotic areas—few to no melanocytes	Defect in migration of melanoblasts from neural crest to involved skin or failure of melanoblasts to survive or differentiate in these areas Mutations within the *KIT* protooncogene that encodes the tyrosine kinase receptor for stem cell growth factor (kit ligand)	None; occasionally transplants

[a]If potassium hydroxide (KOH) examination of scale is negative, consider the possibility of progressive macular hypomelanosis.

Abbreviations: MBEH, monobenzylether of hydroquinone; PUVA, psoralens plus ultraviolet A irradiation; UV-B, ultraviolet B irradiation.

There are four systemic diseases that should be considered in a patient with skin findings suggestive of vitiligo—*systemic sclerosis*, *melanoma-associated leukoderma*, *onchocerciasis*, and *Vogt-Koyanagi-Harada syndrome*. The vitiligo-like leukoderma seen in patients with systemic sclerosis has a clinical resemblance to idiopathic vitiligo that has begun to repigment as a result of treatment; that is, perifollicular macules of normal pigmentation are seen within areas of depigmentation. The basis of this leukoderma is unknown; there is no evidence of inflammation in areas of involvement, but it can resolve if the underlying connective tissue disease becomes inactive. In contrast to idiopathic vitiligo, melanoma-associated vitiligo-like leukoderma often begins on the trunk, and its appearance, if spontaneous, should prompt a search for metastatic disease. It is also seen in patients undergoing immunotherapy for melanoma, including immune checkpoint-blocking antibodies, with cytotoxic T lymphocytes presumably recognizing cell surface antigens common to melanoma cells and melanocytes, and is associated with a greater likelihood of a clinical response. A history of aseptic meningitis, nontraumatic uveitis, tinnitus, hearing loss, and/or dysacousia points to the diagnosis of the Vogt-Koyanagi-Harada syndrome. In these patients, the face and scalp are the most common locations of pigment loss.

There are two systemic disorders (neurocristopathies) that may have the cutaneous findings of piebaldism (Table 58-9). They are *Shah-Waardenburg syndrome* and *Waardenburg syndrome*. A possible explanation for both disorders is an abnormal embryonic migration or survival of two neural crest–derived elements, one of them being melanocytes and the other myenteric ganglion cells (leading to Hirschsprung disease in Shah-Waardenburg syndrome) or auditory nerve cells (Waardenburg syndrome). The latter syndrome is characterized by congenital sensorineural hearing loss, dystopia canthorum (lateral displacement of the inner canthi but normal interpupillary distance), heterochromic irises, and a broad nasal root, in addition to

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

syndrome) or auditory nerve cells (Waardenburg syndrome). The latter syndrome is characterized by congenital sensorineural hearing loss, dystopia canthorum (lateral displacement of the inner canthi but normal interpupillary distance), heterochromic irises, and a broad nasal root, in addition to the piebaldism. The facial dysmorphism can be explained by the neural crest origin of the connective tissues of the head and neck. Patients with Waardenburg syndrome have been shown to have mutations in four genes, including *PAX-3* and *MITF*, all of which encode transcription factors, whereas patients with Hirschsprung disease plus white spotting have mutations in one of three genes—endothelin 3, endothelin B receptor, and *SOX-10*.

In *tuberous sclerosis*, the earliest cutaneous sign is macular hypomelanosis, referred to as an ash leaf spot. These lesions are often present at birth and are usually multiple; however, detection may require Wood's lamp examination, especially in lightly pigmented individuals. The pigment within them is reduced, but not absent. The average size is 1–3 cm, and the common shapes are polygonal and lance-ovate. Examination of the patient for additional cutaneous signs such as multiple angiofibromas of the face (adenoma sebaceum), ungual and intraoral fibromas, fibrous cephalic plaques, and connective tissue nevi (shagreen patches) is recommended. It is important to remember that an ash leaf spot on the scalp will result in a circumscribed patch of lightly pigmented hair. Internal manifestations include seizures, intellectual disability, central nervous system (CNS) and retinal hamartomas, pulmonary lymphangioleiomyomatosis (women), renal angiomyolipomas, and cardiac rhabdomyomas. The latter can be detected in up to 60% of children (<18 years) with tuberous sclerosis by echocardiography.

Nevus depigmentosus is a stable, well-circumscribed hypomelanosis that is present at birth. There is usually a single oval or rectangular lesion, but when there are multiple lesions, the possibility of tuberous sclerosis needs to be considered. In *linear nevoid hypopigmentation*, a term that is replacing hypomelanosis of Ito and segmental or systematized nevus depigmentosus, streaks and swirls of hypopigmentation are observed. Up to one-third of patients in a tertiary care setting had associated abnormalities involving the musculoskeletal system (asymmetry), the CNS (seizures and intellectual disability), and the eyes (strabismus and hypertelorism). Chromosomal mosaicism has been detected in these patients, lending support to the hypothesis that the cutaneous pattern is the result of the migration of two clones of primordial melanocytes, each with a different pigment potential.

Localized areas of decreased pigmentation are commonly seen as a result of cutaneous inflammation (Table 58-10) and have been observed in the skin overlying active lesions of sarcoidosis (see "Papulonodular Skin Lesions," below) as well as in CTCL. Cutaneous infections also present as disorders of hypopigmentation, and in *tuberculoid leprosy*, there are a few asymmetric patches of hypomelanosis that have associated anesthesia, anhidrosis, and alopecia. Biopsy specimens of the palpable border show dermal granulomas that contain rare, if any, *Mycobacterium leprae* organisms.

HYPERPIGMENTATION

(Table 58-11) Disorders of hyperpigmentation are also divided into two major groups—localized and diffuse. The localized forms are due to an epidermal alteration, a proliferation of melanocytes, or an increase in pigment production. Both acanthosis nigricans and seborrheic keratoses belong to the first group. *Acanthosis nigricans* can be a reflection of an internal malignancy, most commonly of the gastrointestinal tract, and it appears as velvety hyperpigmentation, primarily in flexural areas. However, in the majority of patients, acanthosis nigricans is associated with obesity and insulin resistance, although it may be a reflection of an endocrinopathy such as acromegaly, Cushing's syndrome, polycystic ovary syndrome, or insulin-resistant diabetes mellitus (type A, type B, and lipodystrophic forms). *Seborrheic keratoses* are common lesions, but in one rare clinical setting, they are a sign of systemic disease, and that setting is the sudden appearance of multiple lesions, often with an inflammatory base and in association with acrochordons (skin tags) and acanthosis nigricans. This is termed the *sign of Leser-Trélat* and alerts the clinician to search for an internal malignancy.

TABLE 58-11

Causes of Hyperpigmentation

I. Primary cutaneous disorders
  A. Localized
    1. Epidermal alteration
      a. Seborrheic keratosis
      b. Pigmented actinic keratosis
    2. Proliferation of melanocytes
      a. Lentigo
      b. Melanocytic nevus (mole)
      c. Melanoma

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

      c. Postinflammatory hyperpigmentation (also dermal)

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

b. Melanocytic nevus (mole)
c. Melanoma

3. Increased pigment production
   a. Ephelide (freckle)
   b. Café au lait macule
   c. Postinflammatory hyperpigmentation (also dermal)
   d. Melasma (also dermal)
4. Dermal pigmentation
   a. Fixed drug eruption

B. Localized and diffuse
1. Drugs (e.g., minocycline, hydroxychloroquine, bleomycin)

II. Systemic diseases
A. Localized
1. Epidermal alteration
   a. Acanthosis nigricans (insulin resistance > other endocrine disorders, paraneoplastic)
   b. Seborrheic keratoses (sign of Leser-Trélat)
2. Proliferation of melanocytes
   a. Lentigines (Peutz-Jeghers and LEOPARD/Noonan with multiple lentigines syndromes; xeroderma pigmentosum)
   b. Melanocytic nevi (Carney complex [LAMB and NAME syndromes])[a]
3. Increased pigment production
   a. Café au lait macules (neurofibromatosis, Legius syndrome, McCune-Albright syndrome[b])
   b. Urticaria pigmentosa[c]
4. Dermal pigmentation
   a. Incontinentia pigmenti (stage III)
   b. Dyskeratosis congenita
5. Dermal deposits
   a. Exogenous ochronosis
   b. Localized argyria

B. Diffuse
1. Endocrinopathies
   a. Addison's disease
   b. Nelson syndrome
   c. Ectopic ACTH syndrome
   d. Hyperthyroidism
2. Metabolic
   a. Porphyria cutanea tarda
   b. Hemochromatosis
   c. Vitamin B_{12}, folate deficiency
   d. Pellagra
   e. Malabsorption, including Whipple's disease
3. Melanosis secondary to metastatic melanoma
4. Autoimmune
   a. Primary biliary cholangitis
   b. Systemic sclerosis (scleroderma)
   c. POEMS syndrome
   d. Eosinophilia-myalgia syndrome[d]
5. Drugs (e.g., cyclophosphamide) and metals (e.g., silver)

[a]Also lentigines. [b]Polyostotic fibrous dysplasia. [c]See also "Papulonodular Skin Lesions." [d]Late 1980s.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

neurofibroma, and *e*phelides (freckles); POEMS, *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathies, *M*-protein, and *s*kin changes.

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

[a]Also lentigines. [b]Polyostotic fibrous dysplasia. [c]See also "Papulonodular Skin Lesions." [d]Late 1980s.

Abbreviations: LAMB, *l*entigines, *a*trial *m*yxomas, *m*ucocutaneous myxomas, and *b*lue nevi; LEOPARD, *l*entigines, *E*CG abnormalities, *o*cular hypertelorism, *p*ulmonary stenosis and subaortic valvular stenosis, *a*bnormal genitalia, *r*etardation of growth, and *d*eafness (sensorineural); NAME, *n*evi, *a*trial myxoma, *m*yxoid neurofibroma, and *e*phelides (freckles); POEMS, *p*olyneuropathy, *o*rganomegaly, *e*ndocrinopathies, *M*-protein, and *s*kin changes.

A proliferation of melanocytes results in the following pigmented lesions: *lentigo*, *melanocytic nevus*, and *melanoma* **(Chap. 76)**. In an adult, the majority of lentigines are related to sun exposure, which explains their distribution. However, in the Peutz-Jeghers and LEOPARD (*l*entigines; *E*CG abnormalities, primarily conduction defects; *o*cular hypertelorism; *p*ulmonary stenosis and subaortic valvular stenosis; *a*bnormal genitalia [cryptorchidism, hypospadias]; *r*etardation of growth; and *d*eafness [sensorineural]) syndromes, lentigines do serve as a clue to systemic disease. In *LEOPARD/Noonan with multiple lentigines syndrome*, hundreds of lentigines develop during childhood and are scattered over the entire surface of the body. The lentigines in patients with *Peutz-Jeghers syndrome* are located primarily around the nose and mouth, on the hands and feet, and within the oral cavity. While the pigmented macules on the face may fade with age, the oral lesions persist. However, similar intraoral lesions are also seen in Addison's disease, in Laugier-Hunziker syndrome (no internal manifestations), and as a normal finding in darkly pigmented individuals. Patients with this autosomal dominant syndrome (due to mutations in a novel serine threonine kinase gene) have multiple benign polyps of the gastrointestinal tract, testicular or ovarian tumors, and an increased risk of developing gastrointestinal (primarily colon) and pancreatic cancers.

In the *Carney complex,* numerous lentigines are also seen, but they are in association with cardiac myxomas. This autosomal dominant disorder is also known as the *LAMB* (*l*entigines, *a*trial myxomas, *m*ucocutaneous myxomas, and *b*lue nevi) *syndrome* or *NAME* (*n*evi, *a*trial myxoma, *m*yxoid neurofibroma, and *e*phelides [freckles]) *syndrome*. These patients can also have evidence of endocrine overactivity in the form of Cushing's syndrome (pigmented nodular adrenocortical disease) and acromegaly.

The third type of localized hyperpigmentation is due to a local increase in pigment production, and it includes *ephelides* and *café au lait macules* (CALMs). While a single CALM can be seen in up to 10% of the normal population, the presence of multiple or large-sized CALMs raises the possibility of an associated genodermatosis, for example, neurofibromatosis (NF) or McCune-Albright syndrome. *CALMs* are flat, uniformly brown in color (usually two shades darker than uninvolved skin), and can vary in size from 0.5 to 12+ cm. More than 90% of adult patients with *type I NF* will have six or more CALMs measuring ≥1.5 cm in diameter. Additional findings are discussed in the section on neurofibromas (see "Papulonodular Skin Lesions," below). In comparison with NF, the CALMs in patients with *McCune-Albright syndrome* (polyostotic fibrous dysplasia with precocious puberty in females due to mosaicism for an activating mutation in a G protein [$G_s\alpha$] gene) are usually larger, are more irregular in outline, and tend to respect the midline.

In incontinentia pigmenti, dyskeratosis congenita, and bleomycin pigmentation, the areas of localized hyperpigmentation form a pattern—swirls and streaks in the first, reticulated in the second, and flagellate in the third. In *dyskeratosis congenita*, atrophic reticulated hyperpigmentation is seen on the neck, trunk, and thighs and is accompanied by nail dystrophy, pancytopenia, and leukoplakia of the oral and anal mucosae. The latter often develops into squamous cell carcinoma. In addition to the flagellate pigmentation (linear streaks) on the trunk, patients receiving bleomycin often have hyperpigmentation overlying the elbows, knees, and small joints of the hand.

Localized hyperpigmentation is seen as a side effect of several other *systemic medications*, including those that produce fixed drug reactions (nonsteroidal anti-inflammatory drugs [NSAIDs], sulfonamides, barbiturates, and tetracyclines) and those that can complex with melanin or iron (antimalarials and minocycline). Fixed drug eruptions recur in the exact same location as circular areas of erythema that can become bullous and then resolve as brown macules. The eruption usually appears within hours of readministration of the offending agent, and common locations include the genitalia, distal extremities, and perioral region. Chloroquine and hydroxychloroquine produce gray-brown to blue-black discoloration of the shins, hard palate, and face, while blue macules (often misdiagnosed as bruises) can be seen on the lower extremities and in sites of inflammation with prolonged minocycline administration. Estrogen in oral contraceptives can induce melasma—symmetric brown patches on the face, especially the cheeks, upper lip, and forehead. Similar changes are seen in pregnancy and in patients receiving phenytoin.

In the diffuse forms of hyperpigmentation, the darkening of the skin may be of equal intensity over the entire body or may be accentuated in sun-exposed areas. The causes of diffuse hyperpigmentation can be divided into four major groups—endocrine, metabolic, autoimmune, and drugs. The endocrinopathies that frequently have associated hyperpigmentation include *Addison's disease*, *Nelson's syndrome*, and *ectopic adrenocorticotropic hormone (ACTH) syndrome*. In these diseases, the increased pigmentation is diffuse but is accentuated in sun-exposed areas, as well as in the palmar creases, sites of friction, and scars. An overproduction of the pituitary hormones α-MSH (melanocyte-stimulating hormone) and ACTH can lead to an increase in melanocyte activity. These peptides are products of the proopiomelanocortin gene and exhibit homology, for example, α-MSH and ACTH share 13 amino acids. A minority of patients with Cushing's disease or hyperthyroidism have generalized hyperpigmentation.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

exposed areas and is a reflection of the photoreactive properties of porphyrins. The increased level of iron in the skin of patients with type 1 hemochromatosis stimulates melanin pigment production and leads to the classic bronze color. Patients with pellagra have a brown discoloration of

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

increase in melanocyte activity. These peptides are products of the proopiomelanocortin gene and exhibit homology, for example, α-MSH and ACTH share 13 amino acids. A minority of patients with Cushing's disease or hyperthyroidism have generalized hyperpigmentation.

The metabolic causes of hyperpigmentation include *porphyria cutanea tarda* (PCT), *hemochromatosis*, *vitamin B_{12} deficiency*, *folic acid deficiency*, *pellagra*, and *malabsorption*, including *Whipple's disease*. In patients with PCT (see "Vesicles/Bullae," below), the skin darkening is seen in sun-exposed areas and is a reflection of the photoreactive properties of porphyrins. The increased level of iron in the skin of patients with type 1 hemochromatosis stimulates melanin pigment production and leads to the classic bronze color. Patients with pellagra have a brown discoloration of the skin, especially in sun-exposed areas, as a result of nicotinic acid (niacin) deficiency. In the areas of increased pigmentation, there is a thin, varnish-like scale. These changes are also seen in patients who are vitamin B_6 deficient, have functioning carcinoid tumors (increased consumption of niacin), or take isoniazid. Approximately 50% of the patients with Whipple's disease have an associated generalized hyperpigmentation in association with diarrhea, weight loss, arthritis, and lymphadenopathy. A diffuse, slate-blue to gray-brown color is seen in patients with *melanosis secondary to metastatic melanoma*. The color reflects widespread deposition of melanin within the dermis as a result of the high concentration of circulating melanin precursors.

Of the autoimmune diseases associated with diffuse hyperpigmentation, *primary biliary cholangitis* and *systemic sclerosis* are the most common, and occasionally, both disorders are seen in the same patient. The skin is dark brown in color, especially in sun-exposed areas. In primary biliary cholangitis, the hyperpigmentation is accompanied by pruritus, jaundice, and xanthomas, whereas in systemic sclerosis, it is accompanied by sclerosis of the extremities, face, and, less commonly, the trunk. Additional clues to the diagnosis of systemic sclerosis are mat and cuticular telangiectasias, calcinosis cutis, Raynaud's phenomenon, and distal ulcerations (see "Telangiectasias," above). The differential diagnosis of cutaneous sclerosis with hyperpigmentation includes POEMS (*p*olyneuropathy; *o*rganomegaly [liver, spleen, lymph nodes]; *e*ndocrinopathies [impotence, gynecomastia]; *M*-protein; and *s*kin changes) syndrome. The skin changes include hyperpigmentation, induration, hypertrichosis, angiomas, clubbing, and facial lipoatrophy.

Diffuse hyperpigmentation that is due to drugs or metals can result from one of several mechanisms—induction of melanin pigment formation, complexing of the drug or its metabolites to melanin, and deposits of the drug in the dermis. Busulfan, cyclophosphamide, 5-fluorouracil, and inorganic arsenic induce pigment production. Complexes containing melanin or iron plus the drug or its metabolites are seen in patients receiving minocycline, and a diffuse, brown-gray, muddy appearance within sun-exposed areas may develop, in addition to pigmentation of the mucous membranes, teeth, nails, bones, and thyroid. Administration of amiodarone can result in both a phototoxic eruption (exaggerated sunburn) and/or a slate-gray to violaceous discoloration of sun-exposed skin. Biopsy specimens of the latter show yellow-brown granules in dermal macrophages, which represent intralysosomal accumulations of lipids, amiodarone, and its metabolites. Actual deposits of a particular drug or metal in the skin are seen with silver (argyria), where the skin appears blue-gray in color; gold (chrysiasis), where the skin has a brown to blue-gray color; and clofazimine, where the skin appears reddish brown. The associated pigmentation is accentuated in sun-exposed areas, and discoloration of the eye is seen with gold (sclerae) and clofazimine (conjunctivae).

VESICLES/BULLAE

(Table 58-12) Depending on their size, cutaneous blisters are referred to as *vesicles* (<1 cm) or *bullae* (>1 cm). The primary autoimmune blistering disorders include *pemphigus vulgaris*, *pemphigus foliaceus*, *paraneoplastic pemphigus*, *bullous pemphigoid*, *gestational pemphigoid*, *cicatricial pemphigoid*, *epidermolysis bullosa acquisita*, *linear IgA bullous dermatosis* (*LABD*), and *dermatitis herpetiformis* **(Chap. 59)**.

TABLE 58-12

Causes of Vesicles/Bullae

I. Primary mucocutaneous diseases
   A. Primary blistering diseases (autoimmune)
      1. Pemphigus, foliaceus and vulgaris[a]
      2. Bullous pemphigoid[b]
      3. Gestational pemphigoid[b]
      4. Cicatricial pemphigoid[b]
      5. Dermatitis herpetiformis[b,c]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

      1. Contact dermatitis[a,b]
      2. Erythema multiforme[e]

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

I. Primary mucocutaneous diseases
  A. Primary blistering diseases (autoimmune)
    1. Pemphigus, foliaceus and vulgaris[a]
    2. Bullous pemphigoid[b]
    3. Gestational pemphigoid[b]
    4. Cicatricial pemphigoid[b]
    5. Dermatitis herpetiformis[b,c]
    6. Linear IgA bullous dermatosis[b]
    7. Epidermolysis bullosa acquisita[b,d]
  B. Secondary blistering diseases
    1. Contact dermatitis[a,b]
    2. Erythema multiforme[e]
    3. Stevens-Johnson syndrome[e]
    4. Toxic epidermal necrolysis[e]
    5. Bullous fixed drug eruption, including generalized variant[e]
    6. Pseudoporphyria, drug- or tanning booth–induced
  C. Infections
    1. Varicella-zoster virus[a,f]
    2. Herpes simplex virus[a,f]
    3. Enteroviruses, e.g., hand-foot-and-mouth disease[f]
    4. SARS-CoV-2
    Staphylococcal scalded-skin syndrome[a,g]
    5. Bullous impetigo[a]
    6. Bullous tinea
II. Systemic diseases
  A. Autoimmune
    1. Paraneoplastic pemphigus[a] (bronchiolitis obliterans)
  B. Infections
    1. Cutaneous emboli[b]
  C. Metabolic
    1. Diabetic bullae[a,b]
    2. Porphyria cutanea tarda[b]
    3. Porphyria variegata[b]
    4. Bullous dermatosis of hemodialysis[b] (less often associated with peritoneal dialysis and also referred to as pseudoporphyria)
  D. Ischemia
    1. Coma bullae
  E. Secondary blistering diseases
    1. Toxic epidermal necrolysis[e] (respiratory and gastrointestinal tracts can be involved)

[a]Intraepidermal. [b]Subepidermal. [c]Associated with gluten enteropathy. [d]Associated with inflammatory bowel disease. [e]Degeneration of cells within the basal layer of the epidermis can give impression split is subepidermal. [f]Also systemic. [g]In adults, associated with renal failure and immunocompromised state.

Vesicles and bullae are also seen in *contact dermatitis*, both allergic and irritant forms **(Chap. 57)**. When there is a linear arrangement of vesicular

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

are doxycycline, quinolones, voriconazole, thiazides, NSAIDs, vemurafenib, and psoralens. The development of a phototoxic eruption is dependent on the doses of both the drug and ultraviolet (UV)-A irradiation.

Vesicles and bullae are also seen in *contact dermatitis*, both allergic and irritant forms **(Chap. 57)**. When there is a linear arrangement of vesicular lesions, an exogenous cause or herpes zoster should be suspected. Bullous disease secondary to the ingestion of drugs can take one of several forms, including phototoxic eruptions, isolated bullae, Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN) **(Chap. 60)**. Clinically, phototoxic eruptions resemble an exaggerated sunburn with diffuse erythema and bullae in sun-exposed areas. The most commonly associated drugs are doxycycline, quinolones, voriconazole, thiazides, NSAIDs, vemurafenib, and psoralens. The development of a phototoxic eruption is dependent on the doses of both the drug and ultraviolet (UV)-A irradiation.

Toxic epidermal necrolysis is characterized by bullae that arise on widespread areas of tender erythema and then slough. This results in large areas of denuded skin. The associated morbidity, such as sepsis, and mortality rates are relatively high and are a function of the extent of epidermal necrosis. In addition, these patients may also have involvement of the mucous membranes and respiratory and intestinal tracts. Drugs are the primary cause of TEN, and the most common offenders are aromatic anticonvulsants (phenytoin, barbiturates, carbamazepine), sulfonamides, aminopenicillins, allopurinol, and NSAIDs. Generalized bullous fixed drug eruption, severe acute graft-versus-host disease (grade 4), vancomycin-induced LABD, and flares of cutaneous lupus can also resemble TEN.

In *erythema multiforme* (EM), the primary lesions are pink-red macules and edematous papules, the centers of which may become vesicular. In contrast to a morbilliform exanthem, the clue to the diagnosis of EM, and especially SJS, is the development of a "dusky" violet color in the center of the lesions. Target lesions are also characteristic of EM and arise as a result of active centers and borders in combination with centrifugal spread. However, target lesions need not be present to make the diagnosis of EM.

EM has been subdivided into two major groups: (1) EM minor due to herpes simplex virus (HSV); and (2) EM major due to HSV, *Mycoplasma pneumonia*, or, occasionally, other viruses, *Chlamydia*, or drugs. Involvement of the mucous membranes (ocular, nasal, oral, and genital) is seen more commonly in the latter form, and in patients with *Mycoplasma pneumoniae*-induced rash and mucositis (MIRM), there may be minimal cutaneous involvement. Hemorrhagic crusts of the lips are characteristic of EM major and SJS as well as herpes simplex, pemphigus vulgaris, and paraneoplastic pemphigus. Fever, malaise, myalgias, sore throat, and cough may precede or accompany the eruption. The lesions of EM usually resolve over 2–4 weeks but may be recurrent, especially when due to HSV. In addition to HSV (in which lesions usually appear 7–12 days after the viral eruption), EM can also follow vaccinations, radiation therapy, and exposure to environmental toxins, including the oleoresin in poison ivy.

Induction of SJS is most often due to drugs, especially sulfonamides, aromatic anticonvulsants, lamotrigine, aminopenicillins, and nonnucleoside reverse transcriptase inhibitors (e.g., nevirapine). Widespread dusky macules and significant mucosal involvement are characteristic of SJS, and the cutaneous lesions may or may not develop epidermal detachment. If the latter occurs, by definition, it is limited to <10% of the body surface area (BSA). Greater involvement leads to the diagnosis of SJS/TEN overlap (10–30% BSA) or TEN (>30% BSA).

In addition to primary blistering disorders and hypersensitivity reactions, bacterial and viral infections can lead to vesicles and bullae. The most common infectious agents are HSV **(Chap. 192)**, varicella-zoster virus **(Chap. 193)**, and *S. aureus* **(Chap. 147)**.

Staphylococcal scalded-skin syndrome (SSSS) and *bullous impetigo* are two blistering disorders associated with staphylococcal (phage group II) infection. In SSSS, the initial findings are redness and tenderness of the central face, neck, trunk, and intertriginous zones. This is followed by short-lived flaccid bullae and a slough or exfoliation of the superficial epidermis. Crusted areas then develop, characteristically around the mouth in a radial pattern. SSSS is distinguished from TEN by the following features: younger age group (primarily infants and toddlers), more superficial site of blister formation, no oral lesions, shorter course, lower morbidity and mortality rates, and an association with staphylococcal exfoliative toxin ("exfoliatin"), not drugs. A rapid diagnosis of SSSS versus TEN can be made by a frozen section of the blister roof or exfoliative cytology of the blister contents. In SSSS, the site of staphylococcal infection is usually extracutaneous (conjunctivitis, rhinorrhea, otitis media, pharyngitis, tonsillitis), and the cutaneous lesions are sterile, whereas in bullous impetigo, the skin lesions are the site of infection. Impetigo is more localized than SSSS and usually presents with honey-colored crusts. Occasionally, superficial purulent blisters also form. *Cutaneous emboli* from gram-negative infections may present as isolated bullae, but the base of the lesion is purpuric or necrotic, and it may develop into an ulcer (see "Purpura," below).

Several metabolic disorders are associated with blister formation, including diabetes mellitus, renal failure, and porphyria. Local hypoxemia secondary to decreased cutaneous blood flow can also produce blisters, which explains the presence of bullae over pressure points in comatose patients (coma bullae). In *diabetes mellitus*, tense bullae with clear sterile viscous fluid arise on normal skin. The lesions can be as large as 6 cm in diameter and are located on the distal extremities. There are several types of porphyria, but the most common form with cutaneous findings is *porphyria cutanea tarda* (PCT). In sun-exposed areas (primarily the hands), the skin is very fragile, with trauma leading to erosions mixed with tense vesicles. These lesions then heal with scarring and formation of milia; the latter are firm, 1- to 2-mm white or yellow papules that represent epidermoid cysts. Associated findings can include hypertrichosis of the lateral malar region (men) or face (women) and, in sun-exposed areas, hyperpigmentation

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The differential diagnosis of PCT includes (1) *porphyria variegata*—the skin signs of PCT plus the systemic findings of acute intermittent porphyria; it

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

secondary to decreased cutaneous blood flow can also produce blisters, which explains the presence of bullae over pressure points in comatose vesicles. These lesions then heal with scarring and formation of milia; the latter are firm, 1- to 2-mm white or yellow papules that represent epidermoid cysts. Associated findings can include hypertrichosis of the lateral malar region (men) or face (women) and, in sun-exposed areas, hyperpigmentation and firm sclerotic plaques. An elevated level of urinary uroporphyrins confirms the diagnosis and is due to a decrease in uroporphyrinogen decarboxylase activity. PCT can be exacerbated by alcohol, hemochromatosis and other forms of iron overload, chlorinated hydrocarbons, hepatitis C virus and HIV infections, and hepatomas.

The differential diagnosis of PCT includes (1) *porphyria variegata*—the skin signs of PCT plus the systemic findings of acute intermittent porphyria; it has a diagnostic plasma porphyrin fluorescence emission at 626 nm; (2) *drug-induced pseudoporphyria*—the clinical and histologic findings are similar to PCT, but porphyrins are normal; etiologic agents include naproxen and other NSAIDs, furosemide, tetracycline, and voriconazole; (3) *bullous dermatosis of hemodialysis*—the same appearance as PCT, but porphyrins are usually normal or occasionally borderline elevated; patients have chronic renal failure and are on hemodialysis; (4) *PCT associated with hepatomas and hemodialysis;* and (5) *epidermolysis bullosa acquisita* **(Chap. 59)**.

EXANTHEMS

(Table 58-13) Exanthems are characterized by an acute generalized eruption. The most common presentation is erythematous macules and papules (morbilliform) and less often confluent blanching erythema (scarlatiniform). *Morbilliform* eruptions are usually due to either drugs or viral infections. For example, up to 5% of patients receiving penicillins, sulfonamides, phenytoin, or nevirapine will develop a maculopapular eruption. Accompanying signs may include pruritus, fever, eosinophilia, transaminitis, and transient lymphadenopathy **(Chap. 60)**. Similar maculopapular eruptions are seen in the classic childhood viral exanthems, including (1) *rubeola* (measles)—a prodrome of coryza, cough, and conjunctivitis followed by Koplik's spots on the buccal mucosa; the eruption begins behind the ears, at the hairline, and on the forehead and then spreads down the body, often becoming confluent; (2) *rubella*—the eruption begins on the forehead and face and then spreads down the body; it resolves in the same order and is associated with retroauricular and suboccipital lymphadenopathy; and (3) *erythema infectiosum* (fifth disease)—erythema of the cheeks is followed by a reticulated pattern on the extremities; it is secondary to a parvovirus B19 infection, and an associated arthritis is seen in adults.

TABLE 58-13

Causes of Exanthems

I. Morbilliform
   A. Drugs
   B. Viral
      1. Rubeola (measles)
      2. Rubella
      3. Erythema infectiosum (erythema of cheeks; reticulated on extremities)
      4. Epstein-Barr virus, echovirus, coxsackievirus, CMV, adenovirus, HHV-6/HHV-7[a], dengue, Zika, chikungunya, SARS-CoV-2, and West Nile virus infections
      5. HIV seroconversion exanthem (plus mucosal ulcerations)
   C. Bacterial
      1. Typhoid fever
      2. Early secondary syphilis
      3. Early *Rickettsia* infections
      4. Early meningococcemia
      5. Ehrlichiosis
   D. Acute graft-versus-host disease
   E. Kawasaki disease
II. Scarlatiniform
   A. Scarlet fever
   B. Toxic shock syndrome
   C. Kawasaki disease
   D. Early staphylococcal scalded-skin syndrome

Abbreviations: CMV, cytomegalovirus; HHV, human herpesvirus; HIV, human immunodeficiency virus.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 58-13

Causes of Exanthems

I. Morbilliform
- A. Drugs
- B. Viral
  1. Rubeola (measles)
  2. Rubella
  3. Erythema infectiosum (erythema of cheeks; reticulated on extremities)
  4. Epstein-Barr virus, echovirus, coxsackievirus, CMV, adenovirus, HHV-6/HHV-7[a], dengue, Zika, chikungunya, SARS-CoV-2, and West Nile virus infections
  5. HIV seroconversion exanthem (plus mucosal ulcerations)
- C. Bacterial
  1. Typhoid fever
  2. Early secondary syphilis
  3. Early *Rickettsia* infections
  4. Early meningococcemia
  5. Ehrlichiosis
- D. Acute graft-versus-host disease
- E. Kawasaki disease

II. Scarlatiniform
- A. Scarlet fever
- B. Toxic shock syndrome
- C. Kawasaki disease
- D. Early staphylococcal scalded-skin syndrome

[a]Primary infection in infants and reactivation in the setting of immunosuppression.

Abbreviations: CMV, cytomegalovirus; HHV, human herpesvirus; HIV, human immunodeficiency virus.

Both measles and rubella can occur in unvaccinated adults, and an atypical form of measles is seen in adults immunized with either killed measles vaccine or killed vaccine followed in time by live vaccine. In contrast to classic measles, the eruption of atypical measles begins on the palms, soles, wrists, and ankles, and the lesions may become purpuric. The patient with atypical measles can have pulmonary involvement and be quite ill. Rubelliform and roseoliform eruptions are also associated with *Epstein-Barr virus* (5–15% of patients), *echovirus*, *coxsackievirus*, *cytomegalovirus*, *adenovirus*, SARS-CoV-2, and *dengue*, *chikungunya*, and *West Nile virus* infections. Detection of specific IgM antibodies or fourfold elevations in IgG antibodies often allows the proper diagnosis, but polymerase chain reaction (PCR) is gradually replacing serologic assays. Occasionally, a maculopapular drug eruption is a reflection of an underlying viral infection. For example, ~95% of the patients with infectious mononucleosis who are given ampicillin will develop a rash.

Of note, early in the course of infections with *Rickettsia* and meningococcus, prior to the development of petechiae and purpura, the lesions may be erythematous macules and papules. This is also the case in chickenpox prior to the development of vesicles. Maculopapular eruptions are associated with early *HIV infection*, early secondary *syphilis*, *typhoid fever*, and *acute graft-versus-host disease*. In the last, lesions frequently begin on the dorsal hands and forearms; the macular rose spots of typhoid fever involve primarily the anterior trunk.

The prototypic *scarlatiniform* eruption is seen in *scarlet fever* and is due to an erythrogenic toxin produced by bacteriophage-containing group A β-hemolytic streptococci, most commonly in the setting of pharyngitis. This eruption is characterized by diffuse erythema, which begins on the neck and upper trunk, and red follicular puncta. Additional findings include a white strawberry tongue (white coating with red papillae) followed by a red strawberry tongue (red tongue with red papillae); petechiae of the palate; a facial flush with circumoral pallor; linear petechiae in the antecubital fossae; and desquamation of the involved skin, palms, and soles 5–20 days after onset of the eruption. A similar desquamation of the palms and soles

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

In *toxic shock syndrome*, staphylococcal (phage group I) infections produce an exotoxin (TSST-1) that causes the fever and rash as well as

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

strawberry tongue (red tongue with red papillae); petechiae of the palate; a facial flush with circumoral pallor; linear petechiae in the antecubital fossae; and desquamation of the involved skin, palms, and soles 5–20 days after onset of the eruption. A similar desquamation of the palms and soles is seen in toxic shock syndrome (TSS), in Kawasaki disease, and after severe febrile illnesses. Certain strains of staphylococci also produce an erythrotoxin that leads to the same clinical findings as in streptococcal scarlet fever, except that the anti-streptolysin O or DNase B titers are not elevated.

In *toxic shock syndrome*, staphylococcal (phage group I) infections produce an exotoxin (TSST-1) that causes the fever and rash as well as enterotoxins. Initially, the majority of cases were reported in menstruating women who were using tampons. However, other sites of infection, including wounds and nasal packing, can lead to TSS. The diagnosis of TSS is based on clinical criteria **(Chap. 147)**, and three of these involve mucocutaneous sites (diffuse erythema of the skin, desquamation of the palms and soles 1–2 weeks after onset of illness, and involvement of the mucous membranes). The latter is characterized as hyperemia of the vagina, oropharynx, or conjunctivae. Similar systemic findings have been described in *streptococcal toxic shock syndrome* **(Chap. 148)**, and although an exanthem is seen less often than in TSS due to a staphylococcal infection, the underlying infection is often in the soft tissue (e.g., cellulitis).

The cutaneous eruption in *Kawasaki disease* **(Chap. 363)** is polymorphous, but the two most common forms are morbilliform and scarlatiniform. Additional mucocutaneous findings include bilateral conjunctival injection; erythema and edema of the hands and feet followed by desquamation; and diffuse erythema of the oropharynx, red strawberry tongue, and dry fissured lips. This clinical picture can resemble TSS and scarlet fever, but clues to the diagnosis of Kawasaki disease are cervical lymphadenopathy, cheilitis, and thrombocytosis. The most serious associated systemic finding in this disease is coronary aneurysms secondary to arteritis. Seen primarily in children, SARS-CoV-2-associated multisystem inflammatory syndrome must be distinguished from Kawasaki disease. Scarlatiniform eruptions are also seen in the early phase of SSSS (see "Vesicles/Bullae," above), in young adults with *Arcanobacterium haemolyticum* infection, and as reactions to drugs.

URTICARIA

(Table 58-14) *Urticaria* (hives) are transient lesions that are composed of a central wheal surrounded by an erythematous halo or flare. Individual lesions are round, oval, or figurate and are often pruritic. Acute and chronic urticarias have a wide variety of allergic etiologies and reflect edema in the dermis. Urticarial lesions can also be seen in patients with mastocytosis (urticaria pigmentosa), hypo- or hyperthyroidism, Schnitzler's syndrome, and systemic-onset juvenile idiopathic arthritis (Still's disease). In both juvenile- and adult-onset Still's disease, the lesions coincide with the fever spike, are transient, and are due to dermal infiltrates of neutrophils; the latter is also referred to as neutrophilic urticarial dermatosis.

TABLE 58-14

Causes of Urticaria and Angioedema

I. Primary cutaneous disorders
   A. Acute and chronic urticaria[a]
   B. Physical urticaria
      1. Dermographism
      2. Solar urticaria[b]
      3. Cold urticaria[b]
      4. Cholinergic urticaria[b]
   C. Angioedema (hereditary and acquired)[b,c]
II. Systemic diseases
   A. Urticarial vasculitis
   B. Hepatitis B or C viral infection, SARS-CoV-2 infection
   C. Serum sickness
   D. Angioedema (hereditary and acquired)

[a]A small minority develop anaphylaxis. [b]Also systemic. [c]Acquired angioedema can be idiopathic, associated with a lymphoproliferative disorder, or due to a drug, e.g., angiotensin-converting enzyme (ACE) inhibitors.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

affecting ~5% of the population. *Solar urticaria* characteristically occurs within minutes of sun exposure and is a skin sign of one systemic disease—erythropoietic protoporphyria. In addition to the urticaria, these patients have subtle pitted scarring of the nose and hands. *Cold urticaria* is

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

TABLE 58-14

Causes of Urticaria and Angioedema

I. Primary cutaneous disorders
  A. Acute and chronic urticaria[a]
  B. Physical urticaria
    1. Dermographism
    2. Solar urticaria[b]
    3. Cold urticaria[b]
    4. Cholinergic urticaria[b]
  C. Angioedema (hereditary and acquired)[b,c]
II. Systemic diseases
  A. Urticarial vasculitis
  B. Hepatitis B or C viral infection, SARS-CoV-2 infection
  C. Serum sickness
  D. Angioedema (hereditary and acquired)

[a]A small minority develop anaphylaxis. [b]Also systemic. [c]Acquired angioedema can be idiopathic, associated with a lymphoproliferative disorder, or due to a drug, e.g., angiotensin-converting enzyme (ACE) inhibitors.

The common *physical urticarias* include dermographism, solar urticaria, cold urticaria, and cholinergic urticaria. Patients with *dermographism* exhibit linear wheals following minor pressure or scratching of the skin and may be a contributing factor to pruritic dermatoses. It is a common disorder, affecting ~5% of the population. *Solar urticaria* characteristically occurs within minutes of sun exposure and is a skin sign of one systemic disease—erythropoietic protoporphyria. In addition to the urticaria, these patients have subtle pitted scarring of the nose and hands. *Cold urticaria* is precipitated by exposure to the cold, and therefore, exposed areas are usually affected. In occasional patients, the disease is associated with abnormal circulating proteins—more commonly cryoglobulins and less commonly cryofibrinogens. Additional systemic symptoms include wheezing and syncope, thus explaining the need for these patients to avoid swimming in cold water. Autosomal dominantly inherited cold urticaria is associated with dysfunction of cryopyrin. *Cholinergic urticaria* is precipitated by heat, exercise, or emotion and is characterized by small wheals with relatively large flares. It is occasionally associated with wheezing.

Whereas urticarias are the result of dermal edema, subcutaneous edema leads to the clinical picture of *angioedema*. Sites of involvement include the eyelids, lips, tongue, larynx, and gastrointestinal tract as well as the subcutaneous tissue. Angioedema occurs alone or in combination with urticaria, including urticarial vasculitis and the physical urticarias. Both acquired and hereditary (autosomal dominant) forms of angioedema occur **(Chap. 354)**, and in the latter, urticaria is rarely, if ever, seen.

Urticarial vasculitis is an immune complex disease that may be confused with simple urticaria. In contrast to simple urticaria, individual lesions tend to last longer than 24 h and usually develop central petechiae that can be observed even after the urticarial phase has resolved. The patient may also complain of burning rather than pruritus. On biopsy, there is a leukocytoclastic vasculitis of the small dermal blood vessels. Although urticarial vasculitis may be idiopathic in origin, it can be a reflection of an underlying systemic illness such as lupus erythematosus, Sjögren's syndrome, or hereditary complement deficiency. There is a spectrum of urticarial vasculitis that ranges from purely cutaneous to multisystem involvement. The most common systemic signs and symptoms are arthralgias and/or arthritis, nephritis, and crampy abdominal pain, with asthma and chronic obstructive lung disease seen less often. Hypocomplementemia occurs in one- to two-thirds of patients, even in the idiopathic cases. Urticarial vasculitis can also be seen in patients with *hepatitis B* and *hepatitis C* infections and *serum sickness*, but is usually not seen in *serum sickness–like illnesses* (e.g., due to cefaclor, minocycline).

PAPULONODULAR SKIN LESIONS

(Table 58-15) In the *papulonodular diseases*, the lesions are elevated above the surface of the skin and may coalesce to form larger plaques. The

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Papulonodular Skin Lesions According to Color Groups

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

(Table 58-15) In the *papulonodular diseases*, the lesions are elevated above the surface of the skin and may coalesce to form larger plaques. The location, consistency, and color of the lesions are the keys to their diagnosis; this section is organized on the basis of color.

TABLE 58-15

Papulonodular Skin Lesions According to Color Groups

I. White
- A. Calcinosis cutis
- B. Osteoma cutis (also skin-colored or blue)

II. Skin-colored
- A. Rheumatoid nodules
- B. Neurofibromas (von Recklinghausen's disease [NF1])
- C. Angiofibromas (tuberous sclerosis, MEN syndrome, type 1; also pink-red)
- D. Neuromas (MEN syndrome, type 2b)
- E. Adnexal tumors
  1. Basal cell carcinomas (basal cell nevus syndrome)
  2. Tricholemmomas (Cowden disease)
  3. Fibrofolliculomas (Birt-Hogg-Dubé syndrome)
- F. Osteomas (arise in skull and jaw in Gardner syndrome)
- G. Primary cutaneous disorders
  1. Epidermal inclusion cysts[a]
  2. Lipomas

III. Pink/translucent[b]
- A. Amyloidosis, primary systemic
- B. Papular mucinosis/scleromyxedema
- C. Multicentric reticulohistiocytosis

IV. Yellow
- A. Xanthomas
- B. Tophi
- C. Necrobiosis lipoidica
- D. Pseudoxanthoma elasticum
- E. Sebaceous adenomas (Muir-Torre syndrome)

V. Red[b]
- A. Papules
  1. Angiokeratomas (Fabry disease and related lysosomal storage diseases)[c]
  2. Bacillary angiomatosis (primarily in AIDS)
- B. Papules/plaques
  1. Cutaneous lupus erythematosus
  2. Lymphoma cutis
  3. Leukemia cutis
  4. Sweet syndrome
- C. Nodules
  1. Panniculitis
  2. Medium-sized vessel vasculitis (e.g., cutaneous polyarteritis nodosa)
- D. Primary cutaneous disorders
  1. Arthropod bites
  2. Cherry hemangiomas
  3. Infections, e.g., streptococcal cellulitis, sporotrichosis
  4. Polymorphous light eruption

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

- A. Sarcoidosis
- B. Urticaria pigmentosa

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

2. Medium-sized vessel vasculitis (e.g., cutaneous polyarteritis nodosa)
3. Infections, e.g., streptococcal cellulitis, sporotrichosis
4. Polymorphous light eruption
5. Cutaneous lymphoid hyperplasia (lymphocytoma cutis, pseudolymphoma)

VI. Red-brown[b]
- A. Sarcoidosis
- B. Urticaria pigmentosa
- C. Erythema elevatum diutinum (chronic leukocytoclastic vasculitis)
- D. Lupus vulgaris

VII. Blue[b]
- A. Venous malformations (e.g., blue rubber bleb syndrome)
- B. Primary cutaneous disorders
  1. Venous lake
  2. Blue nevus

VIII. Violaceous
- A. Lupus pernio (sarcoidosis)
- B. Lymphoma cutis
- C. Cutaneous lupus erythematosus

IX. Purple
- A. Kaposi's sarcoma, acral angiodermatitis (pseudo-Kaposi's sarcoma)
- B. Angiosarcoma
- C. Palpable purpura (see Table 58-16)
- D. Primary cutaneous disorders
  1. Angiokeratomas of the scrotum and vulva

X. Brown-black[d]

XI. Any color
- A. Metastases

[a]If multiple with childhood onset, consider Gardner syndrome. [b]May have darker hue in more darkly pigmented individuals. [c]More widespread, especially lower trunk and girdle region, and often red-purple in color. [d]See also "Hyperpigmentation."

Abbreviations: MEN, multiple endocrine neoplasia; NF1, neurofibromatosis type 1.

WHITE LESIONS

In *calcinosis cutis*, there are firm white to white-yellow papules with an irregular surface. When the contents are expressed, a chalky white material is seen. *Dystrophic calcification* is seen at sites of previous inflammation or damage to the skin. It develops in acne scars as well as on the distal extremities of patients with systemic sclerosis and in the subcutaneous tissue and intermuscular fascial planes in DM. The latter is more extensive and is more commonly seen in children. An elevated calcium phosphate product, most commonly due to secondary hyperparathyroidism in the setting of renal failure, can lead to nodules of *metastatic calcinosis cutis*, which tend to be subcutaneous and periarticular. These patients can also develop calcification of muscular arteries and subsequent ischemic necrosis (calciphylaxis). *Osteoma cutis*, in the form of small papules, most commonly occurs on the face of individuals with a history of acne vulgaris, whereas plate-like lesions occur in rare genetic syndromes.

SKIN-COLORED LESIONS

There are several types of skin-colored lesions, including epidermoid cysts, lipomas, rheumatoid nodules, neurofibromas, angiofibromas, neuromas, and adnexal tumors such as tricholemmomas. Both *epidermoid cysts* and *lipomas* are very common mobile subcutaneous nodules—the former are rubbery and drain cheeselike material (sebum and keratin) if incised. Lipomas are firm and somewhat lobulated on palpation. When extensive facial epidermoid cysts develop during childhood or there is a family history of such lesions, the patient should be examined for other signs of Gardner syndrome, including osteomas and desmoid tumors. *Rheumatoid nodules* are firm 0.5- to 4-cm nodules that favor the extensor aspect of joints,

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Neurofibromas (benign Schwann cell tumors) are soft papules or nodules that exhibit the "button-hole" sign, that is, they invaginate into the skin with pressure in a manner similar to a hernia. Single lesions are seen in normal individuals, but multiple neurofibromas, usually in combination with six or

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

syndrome, including osteomas and desmoid tumors. *Rheumatoid nodules* are firm 0.5- to 4-cm nodules that favor the extensor aspect of joints, especially the elbows. They are seen in ~20% of patients with rheumatoid arthritis and 6% of patients with Still's disease. Biopsies of the nodules show palisading granulomas. Similar lesions that are smaller and shorter-lived are seen in rheumatic fever.

Neurofibromas (benign Schwann cell tumors) are soft papules or nodules that exhibit the "button-hole" sign; that is, they invaginate into the skin with pressure in a manner similar to a hernia. Single lesions are seen in normal individuals, but multiple neurofibromas, usually in combination with six or more CALMs measuring >1.5 cm (see "Hyperpigmentation," above), axillary freckling, and multiple Lisch nodules, are seen in von Recklinghausen's disease (NF type I) **(Chap. 90)**. In some patients, the neurofibromas are localized and unilateral due to somatic mosaicism.

Angiofibromas are firm pink-red to skin-colored papules that measure from 3 mm to 1.5 cm in diameter. When multiple lesions are located on the central cheeks (adenoma sebaceum), the patient has tuberous sclerosis or multiple endocrine neoplasia (MEN) syndrome, type 1. The former is an autosomal disorder due to mutations in two different genes, and the associated findings are discussed in the section on ash leaf spots as well as in **Chap. 90**.

Neuromas (benign proliferations of nerve fibers) are also firm, skin-colored papules. They are more commonly found at sites of amputations and in rudimentary polydactyly. However, when there are multiple neuromas on the eyelids, lips, distal tongue, and/or oral mucosa, the patient should be investigated for other signs of MEN syndrome, type 2b. Associated findings include marfanoid habitus, protuberant lips, intestinal ganglioneuromas, and medullary thyroid carcinoma (>75% of patients; **Chap. 388**).

Adnexal tumors are derived from pluripotent cells of the epidermis that can differentiate toward hair, sebaceous, or apocrine or eccrine glands, or remain undifferentiated. *Basal cell carcinomas* (BCCs) are examples of adnexal tumors that have little or no evidence of differentiation. Clinically, they are translucent papules with rolled borders, telangiectasias, and central erosion. BCCs commonly arise in sun-damaged skin of the head and neck as well as the upper trunk. When a patient has multiple BCCs, especially prior to age 30, the possibility of the basal cell nevus syndrome should be raised. It is inherited as an autosomal dominant trait and is associated with jaw cysts, palmar and plantar pits, frontal bossing, medulloblastomas, and calcification of the falx cerebri and diaphragma sellae. *Tricholemmomas* are also skin-colored adnexal tumors but differentiate toward hair follicles and can have a wartlike appearance. The presence of multiple tricholemmomas on the face and cobblestoning of the oral mucosa points to the diagnosis of Cowden disease (multiple hamartoma syndrome) due to mutations in the phosphatase and tensin homolog (*PTEN*) gene. Internal organ involvement (in decreasing order of frequency) includes fibrocystic disease and carcinoma of the breast, adenomas and carcinomas of the thyroid, and gastrointestinal polyposis. Keratoses of the palms, soles, and dorsal aspect of the hands are also seen. *Fibrofolliculomas* are skin-colored to white, smooth papules that favor the face, ears, and neck and, when multiple, are associated with Birt-Hogg-Dubé syndrome, which is associated with renal lesions including cancer **(Chap. 85)**.

PINK LESIONS

The cutaneous lesions associated with primary systemic *amyloidosis* are often pink to pink-orange in color and translucent. Common locations are the face, especially the periorbital and perioral regions, and flexural areas. On biopsy, homogeneous deposits of amyloid are seen in the dermis and in the walls of blood vessels; the latter lead to an increase in vessel wall fragility. As a result, petechiae and purpura develop in clinically normal skin as well as in lesional skin following minor trauma, hence the term *pinch purpura*. Amyloid deposits are also seen in the striated muscle of the tongue and result in macroglossia.

Even though specific mucocutaneous lesions are present in only ~30% of the patients with primary systemic (AL) amyloidosis, the diagnosis can be made via histologic examination of abdominal subcutaneous fat, in conjunction with a serum free light chain assay. By special staining, amyloid deposits are seen around blood vessels or individual fat cells in 40–50% of patients. There are also three forms of amyloidosis that are limited to the skin and that should not be construed as cutaneous lesions of systemic amyloidosis. They are macular amyloidosis (upper back), lichen amyloidosis (usually lower extremities), and nodular amyloidosis. In macular and lichen amyloidosis, the deposits are composed of altered epidermal keratin. Early-onset macular and lichen amyloidosis have been associated with MEN syndrome, type 2a.

Patients with *multicentric reticulohistiocytosis* also have pink-colored papules and nodules on the face and mucous membranes as well as on the extensor surface of the hands and forearms. They have a polyarthritis that can mimic rheumatoid arthritis clinically. On histologic examination, the papules have characteristic giant cells that are not seen in biopsies of rheumatoid nodules. Pink to skin-colored papules that are firm, 2–5 mm in diameter, and often in a linear arrangement are seen in patients with *papular mucinosis*. This disease is also referred to as *scleromyxedema*. The latter name comes from the induration of the face and extremities that may accompany the papular eruption. Biopsy specimens of the papules show localized mucin deposition, and serum protein electrophoresis plus immunofixation electrophoresis demonstrates a monoclonal spike of IgG, usually

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Several systemic disorders are characterized by yellow-colored cutaneous papules or plaques—hyperlipidemia (xanthomas), gout (tophi), diabetes

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

name comes from the induration of the face and extremities that may accompany the papular eruption. Biopsy specimens of the papules show localized mucin deposition, and serum protein electrophoresis plus immunofixation electrophoresis demonstrates a monoclonal spike of IgG, usually with a λ light chain.

YELLOW LESIONS

Several systemic disorders are characterized by yellow-colored cutaneous papules or plaques—hyperlipidemia (xanthomas), gout (tophi), diabetes (necrobiosis lipoidica), pseudoxanthoma elasticum, and Muir-Torre syndrome (sebaceous tumors). Eruptive xanthomas are the most common form of *xanthomas* and are associated with hypertriglyceridemia (primarily hyperlipoproteinemia types I, IV, and V). Crops of yellow papules with erythematous halos occur primarily on the extensor surfaces of the extremities and the buttocks, and they spontaneously involute with a fall in serum triglycerides. Types II and III result in one or more of the following types of xanthoma: xanthelasma, tendon xanthomas, and plane xanthomas. Xanthelasma are found on the eyelids, whereas tendon xanthomas are frequently associated with the Achilles and extensor finger tendons; plane xanthomas are flat and favor the palmar creases and flexural folds. Tuberous xanthomas are frequently associated with hypercholesterolemia; however, they are also seen in patients with hypertriglyceridemia and are found most frequently over the large joints or hand. Biopsy specimens of xanthomas show collections of lipid-containing macrophages (foam cells).

Patients with several disorders, including biliary cirrhosis, can have a secondary form of hyperlipidemia with associated tuberous and plane xanthomas. However, patients with plasma cell dyscrasias have *normolipemic plane xanthomas*. This latter form of xanthoma may be ≥12 cm in diameter and is most frequently seen on the neck, upper trunk, and flexural folds. It is important to note that the most common setting for eruptive xanthomas is uncontrolled diabetes mellitus. The least specific sign for hyperlipidemia is xanthelasma, because at least 50% of the patients with this finding have normal lipid profiles.

In *tophaceous gout*, there are deposits of monosodium urate in the skin around the joints, particularly those of the hands and feet. Additional sites of *tophi* formation include the helix of the ear and the olecranon and prepatellar bursae. The lesions are firm, yellow to yellow-white in color, and occasionally discharge a chalky material. Their size varies from 1 mm to 7 cm, and the diagnosis can be established by polarized light microscopy of the aspirated contents of a tophus. Lesions of *necrobiosis lipoidica* are found primarily on the shins (90%), and patients can have diabetes mellitus or develop it subsequently. Characteristic findings include a central yellow color, atrophy (transparency), telangiectasias, and a red to red-brown border. Ulcerations can also develop within the plaques. Biopsy specimens show necrobiosis of collagen and granulomatous inflammation.

In *pseudoxanthoma elasticum* (PXE), due to mutations in the gene *ABCC6*, there is an abnormal deposition of calcium on the elastic fibers of the skin, eye, and blood vessels. In the skin, the flexural areas such as the neck, axillae, antecubital fossae, and inguinal area are the primary sites of involvement. Yellow papules coalesce to form reticulated plaques that have an appearance similar to that of plucked chicken skin. In severely affected skin, hanging, redundant folds develop. Biopsy specimens of involved skin show swollen and irregularly clumped elastic fibers with deposits of calcium. In the eye, the calcium deposits in Bruch's membrane lead to angioid streaks and choroiditis; in the arteries of the heart, kidney, gastrointestinal tract, and extremities, the deposits lead to angina, hypertension, gastrointestinal bleeding, and claudication, respectively.

Adnexal tumors that have differentiated toward sebaceous glands include sebaceous adenoma, sebaceous carcinoma, and sebaceous hyperplasia. Except for sebaceous hyperplasia, which is commonly seen on the face, these tumors are fairly rare. Patients with Muir-Torre syndrome have one or more *sebaceous adenoma(s)*, and they can also have sebaceous carcinomas and sebaceous hyperplasia as well as keratoacanthomas. The internal manifestations of Muir-Torre syndrome include *multiple* carcinomas of the gastrointestinal tract (primarily colon) as well as cancers of the genitourinary tract.

RED LESIONS

Cutaneous lesions that are red in color have a wide variety of etiologies; in an attempt to simplify their identification, they will be subdivided into papules, papules/plaques, and subcutaneous nodules. Common red papules include *arthropod bites* and *cherry hemangiomas*; the latter are small, bright-red, dome-shaped papules that represent a benign proliferation of capillaries. In patients with AIDS **(Chap. 202)**, the development of multiple red hemangioma-like lesions points to bacillary angiomatosis, and biopsy specimens show clusters of bacilli that stain positively with the Warthin-Starry stain; the pathogens have been identified as *Bartonella henselae* and *Bartonella quintana*. Disseminated visceral disease is seen primarily in immunocompromised hosts but can occur in immunocompetent individuals.

Multiple *angiokeratomas* are seen in Fabry disease, an X-linked recessive lysosomal storage disease that is due to a deficiency of α-galactosidase A. The lesions are red to red-purple in color and can be quite small in size (1–3 mm), with the most common location being the lower trunk. Associated

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

There are several infectious diseases that present as erythematous papules or nodules in a lymphocutaneous or sporotrichoid pattern, that is, in a

lesions are red to red-purple in color and can be quite small in size (1–3 mm), with the most common location being the lower trunk. Associated findings include chronic renal disease, peripheral neuropathy, and corneal opacities (cornea verticillata). While electron photomicrographs demonstrate lamellar lipid deposits in dermal fibroblasts, pericytes, and endothelial cells, nowadays, genetic analysis is more frequently performed for diagnosis. Widespread acute eruptions of erythematous papules are discussed in the section on exanthems.

There are several infectious diseases that present as erythematous papules or nodules in a lymphocutaneous or sporotrichoid pattern, that is, in a linear arrangement along the lymphatic channels. The two most common etiologies are *Sporothrix schenckii* (sporotrichosis) and the atypical mycobacterium *Mycobacterium marinum*. The organisms are introduced as a result of trauma, and a primary inoculation site is often seen in addition to the lymphatic nodules. Additional causes include *Nocardia*, *Leishmania*, and other atypical mycobacteria and dimorphic fungi; culture or PCR of lesional tissue will aid in the diagnosis.

The diseases that are characterized by erythematous plaques with scale are reviewed in the papulosquamous section, and the various forms of dermatitis are discussed in the section on erythroderma. Additional disorders in the differential diagnosis of red papules/plaques include *cellulitis*, *polymorphous light eruption* (PMLE), *cutaneous lymphoid hyperplasia* (lymphocytoma cutis), *cutaneous lupus*, *lymphoma cutis*, and *leukemia cutis*. The first three diseases represent primary cutaneous disorders, although cellulitis may be accompanied by a bacteremia. PMLE is characterized by erythematous papules and plaques in a primarily sun-exposed distribution—dorsum of the hand, extensor forearm, and upper trunk. Lesions follow exposure to UV-B and/or UV-A, and in higher latitudes, PMLE is most severe in the late spring and early summer. A process referred to as "hardening" occurs with continued UV exposure, and the eruption fades, but in temperate climates, it recurs the next spring. PMLE must be differentiated from cutaneous lupus, and this is accomplished by observation of the natural history, histologic examination, and sometimes direct immunofluorescence of the lesions. Cutaneous lymphoid hyperplasia (pseudolymphoma) is a *benign* polyclonal proliferation of lymphocytes within the skin that presents as infiltrated pink-red to red-purple papules and plaques; it must be distinguished from lymphoma cutis.

Several types of red plaques are seen in patients with systemic *lupus*, including (1) erythematous urticarial plaques across the cheeks and nose in the classic butterfly rash; (2) erythematous discoid lesions with fine or "carpet-tack" scale, telangiectasias, central hypopigmentation, peripheral hyperpigmentation, follicular plugging, and atrophy located on the scalp, face, external ears, arms, and upper trunk; and (3) psoriasiform or annular lesions of subacute cutaneous lupus with hypopigmented centers located primarily on the extensor arms and upper trunk. Additional mucocutaneous findings include (1) a violaceous flush on the face and V of the neck; (2) photosensitivity; (3) urticarial vasculitis (see "Urticaria," above); (4) lupus panniculitis (see below); (5) diffuse alopecia; (6) alopecia secondary to discoid lesions; (7) nailfold telangiectasias and erythema; (8) EM- or TEN-like lesions that may become bullous; (9) oral or nasal ulcers; (10) livedo reticularis; and (11) distal ulcerations secondary to Raynaud's phenomenon, vasculitis, or livedoid vasculopathy. Patients with only discoid lesions usually have the form of lupus that is limited to the skin. However, up to 10–15% of these patients eventually develop systemic lupus. Direct immunofluorescence of involved skin, in particular discoid lesions, shows deposits of IgG or IgM and C3 in a granular distribution along the dermal-epidermal junction.

In *lymphoma cutis*, there is a clonal proliferation of malignant lymphocytes within the skin, and the clinical appearance resembles that of cutaneous lymphoid hyperplasia—infiltrated pink-red to red-purple papules and plaques. Lymphoma cutis can occur anywhere on the surface of the skin, whereas the sites of predilection for lymphocytomas include the malar ridge, tip of the nose, and earlobes. Patients with non-Hodgkin's lymphomas have specific cutaneous lesions more often than those with Hodgkin's lymphoma, and, occasionally, the skin nodules precede the development of extracutaneous non-Hodgkin's lymphoma or represent the only site of involvement (e.g., primary cutaneous B-cell lymphoma). Arcuate lesions are sometimes seen in lymphoma and lymphocytoma cutis as well as in CTCL. *Adult T-cell leukemia/lymphoma* that develops in association with HTLV-1 infection is characterized by cutaneous plaques, hypercalcemia, and circulating CD25+ lymphocytes. *Leukemia cutis* has the same appearance as lymphoma cutis, and specific lesions are seen more commonly in monocytic leukemias than in lymphocytic or granulocytic leukemias. Cutaneous chloromas (granulocytic sarcomas) may precede the appearance of circulating blasts in acute myelogenous leukemia and, as such, represent a form of aleukemic leukemia cutis.

Sweet syndrome is characterized by pink-red to red-brown edematous plaques that are frequently painful and occur primarily on the head, neck, and upper extremities. The patients also have fever, neutrophilia, and a dense dermal infiltrate of neutrophils in the lesions. In ~10% of the patients, there is an associated malignancy, most commonly acute myelogenous leukemia. Sweet syndrome has also been reported with inflammatory bowel disease, systemic lupus erythematosus, and solid tumors (primarily of the genitourinary tract) as well as drugs (e.g., granulocyte colony-stimulating factor [G-CSF], hypomethylating agents, all-*trans*-retinoic acid). The differential diagnosis includes neutrophilic eccrine hidradenitis; bullous forms of pyoderma gangrenosum; and, occasionally, cellulitis. Extracutaneous sites of involvement include joints, muscles, eyes, kidneys (proteinuria, occasionally glomerulonephritis), and lungs (neutrophilic infiltrates). The idiopathic form of Sweet syndrome is seen more often in women, following a respiratory tract infection.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

erythema nodosum, erythema induratum/nodular vasculitis, lupus panniculitis, lipodermatosclerosis, α_1-antitrypsin deficiency, factitial, and fat necrosis secondary to pancreatic disease. Except for erythema nodosum, these lesions may break down and ulcerate or heal with a scar. The shin is the

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

tract infection.

Common causes of erythematous subcutaneous nodules include inflamed epidermoid cysts, acne cysts, and furuncles. *Panniculitis*, an inflammation of the fat, also presents as subcutaneous nodules and is frequently a sign of systemic disease. There are several forms of panniculitis, including erythema nodosum, erythema induratum/nodular vasculitis, lupus panniculitis, lipodermatosclerosis, α_1-antitrypsin deficiency, factitial, and fat necrosis secondary to pancreatic disease. Except for erythema nodosum, these lesions may break down and ulcerate or heal with a scar. The shin is the most common location for the nodules of erythema nodosum, whereas the calf is the most common location for lesions of erythema induratum. In erythema nodosum, the nodules are initially red but then develop a blue bruise-like color as they resolve. Patients with erythema nodosum but no underlying systemic illness can still have fever, malaise, leukocytosis, arthralgias, and/or arthritis. However, the possibility of an underlying illness should be excluded, and the most common associations are streptococcal infections, upper respiratory viral infections, sarcoidosis, and inflammatory bowel disease, in addition to drugs (oral contraceptives, sulfonamides, penicillins, bromides, iodides, BRAF inhibitors). Less common associations include bacterial gastroenteritis (*Yersinia*, *Salmonella*) and coccidioidomycosis followed by tuberculosis, histoplasmosis, brucellosis, and infections with *Chlamydia pneumoniae*, *Chlamydia trachomatis, Mycoplasma pneumoniae,* or hepatitis B virus.

Erythema induratum and nodular vasculitis have overlapping features clinically and histologically, and whether they represent two separate entities or the ends of a single disease spectrum is a point of debate; in general, the latter is usually idiopathic and the former is associated with the presence of *Mycobacterium tuberculosis* DNA by PCR within skin lesions. The lesions of lupus panniculitis are found primarily on the cheeks, upper arms, and buttocks (sites of abundant fat) and are seen in both the cutaneous and systemic forms of lupus. The overlying skin may be normal, erythematous, or have the changes of discoid lupus. The subcutaneous fat necrosis that is associated with pancreatic disease is presumably secondary to circulating lipases and is seen in patients with pancreatic carcinoma as well as in patients with acute and chronic pancreatitis. In this disorder, there may be an associated arthritis, fever, and inflammation of visceral fat. Histologic examination of deep incisional biopsy specimens will aid in the diagnosis of the particular type of panniculitis.

Subcutaneous erythematous nodules are also seen in cutaneous polyarteritis nodosa and as a manifestation of *systemic vasculitis* when there is involvement of medium-sized vessels, for example, systemic polyarteritis nodosa, eosinophilic granulomatosis with polyangiitis, or granulomatosis with polyangiitis **(Chap. 363)**. Cutaneous polyarteritis nodosa presents with painful subcutaneous nodules and ulcers within a red-purple, netlike pattern of livedo reticularis. The latter is due to slowed blood flow through the superficial horizontal venous plexus. The majority of lesions are found on the lower extremities, and while arthralgias and myalgias may accompany cutaneous polyarteritis nodosa, there is no evidence of systemic involvement. In both the cutaneous and systemic forms of vasculitis, skin biopsy specimens of the associated nodules will show the changes characteristic of a necrotizing vasculitis and/or granulomatous inflammation.

RED-BROWN LESIONS

The cutaneous lesions in *sarcoidosis* **(Chap. 367)** are classically red to red-brown in color, and with diascopy (pressure with a glass slide), a yellow-brown residual color is observed that is secondary to the granulomatous infiltrate. The waxy papules and plaques may be found anywhere on the skin, but the face is the most common location. Usually there are no surface changes, but occasionally, the lesions will have scale. Biopsy specimens of the papules show “naked” granulomas in the dermis, that is, granulomas surrounded by a minimal number of lymphocytes. Other cutaneous findings in sarcoidosis include annular lesions with an atrophic or scaly center, papules within scars, hypopigmented papules and patches, subcutaneous plaques, alopecia, acquired ichthyosis, erythema nodosum, and lupus pernio (see below).

The differential diagnosis of sarcoidosis includes foreign-body granulomas produced by chemicals such as beryllium and zirconium, late secondary syphilis, and *lupus vulgaris*. Lupus vulgaris is a form of cutaneous tuberculosis that is seen in previously infected and sensitized individuals. There is often underlying active tuberculosis elsewhere, usually in the lungs or lymph nodes. Lesions occur primarily in the head and neck region and are red-brown plaques with a yellow-brown color on diascopy. Secondary scarring can develop within the central portion of the plaques. Cultures or PCR analysis of the lesions should be performed, along with an interferon γ release assay of peripheral blood, because it is rare for the acid-fast stain to show bacilli within the dermal granulomas.

A generalized distribution of red-brown macules and papules is seen in the form of mastocytosis known as *urticaria pigmentosa* **(Chap. 354)**. Each lesion represents a collection of mast cells in the dermis, with hyperpigmentation of the overlying epidermis. Stimuli such as rubbing cause these mast cells to degranulate, and this leads to the formation of localized urticaria (Darier’s sign). Additional symptoms can result from mast cell degranulation and include headache, flushing, diarrhea, and pruritus. Mast cells also infiltrate various organs such as the liver, spleen, and gastrointestinal tract, and accumulations of mast cells in the bones may produce either osteosclerotic or osteolytic lesions on radiographs. In the majority of these patients,

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

BLUE LESIONS

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

accumulations of mast cells in the bones may produce either osteosclerotic or osteolytic lesions on radiographs. In the majority of these patients, however, the internal involvement remains indolent. A subtype of chronic cutaneous small-vessel vasculitis, *erythema elevatum diutinum* (EED), also presents with papules that are red-brown in color. The papules coalesce into plaques on the extensor surfaces of knees, elbows, and the small joints of the hand. Flares of EED have been associated with streptococcal infections.

BLUE LESIONS

Lesions that are blue in color are the result of vascular ectasias, hyperplasias, and tumors or melanin pigment within the dermis. *Venous lakes* (ectasias) are compressible dark-blue lesions that are found commonly in the head and neck region. *Venous malformations* are also compressible blue papulonodules and plaques that can occur anywhere on the body, including the oral mucosa. When there are multiple papulonodules rather than a single congenital lesion, the patient may have the blue rubber bleb syndrome or Maffucci's syndrome. Patients with the blue rubber bleb syndrome also have vascular anomalies of the gastrointestinal tract that may bleed, whereas patients with Maffucci's syndrome have associated osteochondromas. *Blue nevi* (moles) are seen when there are collections of pigment-producing nevus cells in the dermis. These benign papular lesions are dome-shaped and occur most commonly on the dorsum of the hand or foot or in the head and neck region.

VIOLACEOUS LESIONS

Violaceous papules and plaques are seen in *lupus pernio*, *lymphoma cutis*, and *cutaneous lupus*. Lupus pernio is a particular type of sarcoidosis that involves the tip and alar rim of the nose as well as the earlobes, with lesions that are violaceous in color rather than red-brown. This form of sarcoidosis is associated with involvement of the upper respiratory tract. The plaques of lymphoma cutis and cutaneous lupus may be red or violaceous in color and were discussed above.

PURPLE LESIONS

Purple-colored papules and plaques are seen in vascular tumors, such as *Kaposi's sarcoma* (Chap. 202) and *angiosarcoma*, and when there is extravasation of red blood cells into the skin in association with inflammation, as in *palpable purpura* (see "Purpura," below). Patients with congenital or acquired AV fistulas and venous hypertension can develop purple papules on the lower extremities that can resemble Kaposi's sarcoma clinically and histologically; this condition is referred to as pseudo-Kaposi's sarcoma (acral angiodermatitis). Angiosarcoma is found most commonly on the scalp and face of elderly patients or within areas of chronic lymphedema and presents as purple papules and plaques. In the head and neck region, the tumor often extends beyond the clinically defined borders and may be accompanied by facial edema.

BROWN AND BLACK LESIONS

Brown- and black-colored papules are reviewed in "Hyperpigmentation," above.

CUTANEOUS METASTASES

These are discussed last because they can have a wide range of colors. Most commonly, they present as either firm, skin-colored subcutaneous nodules or firm, red to red-brown papulonodules, whereas metastatic melanoma can be pink, blue, or black in color. Cutaneous metastases develop from hematogenous or lymphatic spread and are most often due to the following primary carcinomas: in men, melanoma, oropharynx, lung, and colon; and in women, breast, melanoma, and ovary. These metastatic lesions may be the initial presentation of the carcinoma, especially when the primary site is the lung.

PURPURA

(Table 58-16) *Purpura* are seen when there is an extravasation of red blood cells into the dermis and, as a result, the lesions do not blanch with pressure. This is in contrast to those erythematous or violet-colored lesions that are due to localized vasodilatation—they do blanch with pressure. Purpura (≥3 mm) and petechiae (≤2 mm) are divided into two major groups: palpable and nonpalpable. The most frequent causes of *nonpalpable* purpura and petechiae are primary cutaneous disorders such as *trauma*, *solar* (*actinic*) *purpura*, *stasis purpura*, and *capillaritis*. Less common causes are *steroid purpura* and *livedoid vasculopathy* (see "Ulcers," below). Solar purpura are seen primarily on the extensor forearms, whereas steroid purpura secondary to potent topical glucocorticoids or endogenous or exogenous Cushing's syndrome can be more widespread. In both cases, there is alteration of the supporting connective tissue that surrounds the dermal blood vessels. In contrast, the petechiae that result from capillaritis are found primarily on the lower extremities. In capillaritis, there is an extravasation of erythrocytes as a result of perivascular lymphocytic inflammation.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 58-16

Causes of Purpura

Purpura (≥3 mm) and petechiae (≤2 mm) are divided into two major groups: palpable and nonpalpable. The most frequent causes of *nonpalpable* is alteration of the supporting connective tissue that surrounds the dermal blood vessels. In contrast, the petechiae that result from capillaritis are found primarily on the lower extremities. In capillaritis, there is an extravasation of erythrocytes as a result of perivascular lymphocytic inflammation. The petechiae are bright red, 1–2 mm in size, and scattered within yellow-brown patches. The yellow-brown color is caused by hemosiderin deposits within the dermis.

TABLE 58-16

Causes of Purpura

I. Primary cutaneous disorders
  A. Nonpalpable
    1. Trauma
    2. Solar (actinic, senile) purpura
    3. Steroid purpura
    4. Stasis purpura due to venous hypertension
    5. Capillaritis
    6. Livedoid vasculopathy in the setting of venous hypertension[a]

II. Drugs (e.g., antiplatelet agents, anticoagulants)

III. Systemic diseases
  A. Nonpalpable
    1. Clotting disturbances
      a. Thrombocytopenia (including ITP)
      b. Abnormal platelet function
      c. Clotting factor defects
    2. Vascular fragility
      a. Amyloidosis (within normal-appearing skin)
      b. Ehlers-Danlos syndrome
      c. Scurvy
    3. Thrombi
      a. Disseminated intravascular coagulation, purpura fulminans
      b. Warfarin (Coumadin)-induced necrosis
      c. Heparin-induced thrombocytopenia and thrombosis
      d. Antiphospholipid antibody syndrome
      e. Monoclonal cryoglobulinemia
      f. Vasculopathy induced by levamisole-adulterated cocaine[b]
      g. SARS-CoV-2 infection
      Thrombotic thrombocytopenic purpura
      h. Thrombocytosis
      i. Homozygous protein C or protein S deficiency
    4. Emboli
      a. Cholesterol
      b. Fat
    5. Possible immune complex
      a. Gardner-Diamond syndrome (autoerythrocyte sensitivity)
      b. Waldenström's hypergammaglobulinemic purpura
  B. Palpable
    1. Vasculitis
      a. Cutaneous small-vessel vasculitis, including in the setting of systemic vasculitides
    2. Emboli[c]
      a. Acute meningococcemia
      b. Disseminated gonococcal infection

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

I. Primary cutaneous disorders
  A. Nonpalpable
    1. Trauma
    2. Solar (actinic, senile) purpura
    3. Steroid purpura
    4. Stasis purpura due to venous hypertension
    5. Capillaritis
    6. Livedoid vasculopathy in the setting of venous hypertension[a]
II. Drugs (e.g., antiplatelet agents, anticoagulants)
III. Systemic diseases
  A. Nonpalpable
    1. Clotting disturbances
      a. Thrombocytopenia (including ITP)
      b. Abnormal platelet function
      c. Clotting factor defects
    2. Vascular fragility
      a. Amyloidosis (within normal-appearing skin)
      b. Ehlers-Danlos syndrome
      c. Scurvy
    3. Thrombi
      a. Disseminated intravascular coagulation, purpura fulminans
      b. Warfarin (Coumadin)-induced necrosis
      c. Heparin-induced thrombocytopenia and thrombosis
      d. Antiphospholipid antibody syndrome
      e. Monoclonal cryoglobulinemia
      f. Vasculopathy induced by levamisole-adulterated cocaine[b]
      g. SARS-CoV-2 infection
      Thrombotic thrombocytopenic purpura
      h. Thrombocytosis
      i. Homozygous protein C or protein S deficiency
    4. Emboli
      a. Cholesterol
      b. Fat
    5. Possible immune complex
      a. Gardner-Diamond syndrome (autoerythrocyte sensitivity)
      b. Waldenström's hypergammaglobulinemic purpura
  B. Palpable
    1. Vasculitis
      a. Cutaneous small-vessel vasculitis, including in the setting of systemic vasculitides
    2. Emboli[c]
      a. Acute meningococcemia
      b. Disseminated gonococcal infection
      c. Rocky Mountain spotted fever
      d. Ecthyma gangrenosum

[a]Also associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency. [b]Combined

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Systemic causes of nonpalpable purpura fall into several categories, and those secondary to clotting disturbances and vascular fragility will be discussed first. The former group includes *thrombocytopenia* (Chap. 115), *abnormal platelet function* as is seen in uremia, and *clotting factor*

[a]Also associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency. [b]Combined vasculopathy/vasculitis can be seen. [c]Bacterial (including rickettsial), fungal, or parasitic.

Abbreviation: ITP, idiopathic thrombocytopenic purpura.

Systemic causes of nonpalpable purpura fall into several categories, and those secondary to clotting disturbances and vascular fragility will be discussed first. The former group includes *thrombocytopenia* **(Chap. 115)**, *abnormal platelet function* as is seen in uremia, and *clotting factor defects.* The initial site of presentation for thrombocytopenia-induced petechiae is the distal lower extremity. Capillary fragility leads to nonpalpable purpura in patients with systemic *amyloidosis* (see "Papulonodular Skin Lesions," above), disorders of collagen production such as *Ehlers-Danlos syndrome*, and *scurvy*. In scurvy, there are flattened corkscrew hairs with surrounding hemorrhage on the lower extremities, in addition to gingivitis. Vitamin C is a cofactor for lysyl hydroxylase, an enzyme involved in the posttranslational modification of procollagen that is necessary for cross-link formation.

In contrast to the previous group of disorders, the noninflammatory purpura seen in the following group of diseases are associated with thrombi formation within vessels and have a retiform configuration. It is important to note that these thrombi are demonstrable in skin biopsy specimens. This group of disorders includes disseminated intravascular coagulation (DIC), monoclonal cryoglobulinemia, thrombocytosis, thrombotic thrombocytopenic purpura, antiphospholipid antibody syndrome, and reactions to warfarin and heparin (heparin-induced thrombocytopenia and thrombosis). DIC is triggered by several types of infection (gram-negative, gram-positive, viral, and rickettsial) as well as by tissue injury and neoplasms. Widespread purpura and hemorrhagic infarcts of the distal extremities are seen. Similar lesions are found in purpura fulminans, which is a form of DIC associated with fever and hypotension that occurs more commonly in children following an infectious illness such as varicella, scarlet fever, or an upper respiratory tract infection. In both disorders, hemorrhagic bullae can develop in involved skin.

Monoclonal cryoglobulinemia is associated with plasma cell dyscrasias, chronic lymphocytic leukemia, and lymphoma. Purpura, primarily of the lower extremities, and hemorrhagic infarcts of the fingers, toes, nose and ears are seen in these patients. Exacerbations of disease activity can follow cold exposure or an increase in serum viscosity. Biopsy specimens show precipitates of the cryoglobulin within dermal vessels. Similar deposits have been found in the lung, brain, and renal glomeruli. Patients with *thrombotic thrombocytopenic purpura* can also have hemorrhagic infarcts as a result of intravascular thromboses. Additional signs include microangiopathic hemolytic anemia and fluctuating neurologic abnormalities, especially headaches and confusion.

Administration of *warfarin* can result in painful areas of erythema that become purpuric and then necrotic with an adherent black eschar; the condition is also referred to as Coumadin-induced necrosis. This reaction is seen more often in women and in areas with abundant subcutaneous fat—breasts, abdomen, buttocks, thighs, and calves. The erythema and purpura develop between the third and tenth day of therapy, most likely as a result of a transient imbalance in the levels of anticoagulant and procoagulant vitamin K–dependent factors. Continued therapy does not exacerbate preexisting lesions, and patients with an inherited or acquired deficiency of protein C are at increased risk for this particular reaction as well as for purpura fulminans and calciphylaxis.

Purpura secondary to *cholesterol emboli* are usually seen on the lower extremities of patients with atherosclerotic vascular disease. They often follow anticoagulant therapy or an invasive vascular procedure such as an arteriogram but also occur spontaneously from disintegration of atheromatous plaques. Associated findings include livedo reticularis, gangrene, cyanosis, and ischemic ulcerations. Multiple step sections of the biopsy specimen may be necessary to demonstrate the cholesterol clefts within the vessels. Petechiae are also an important sign of *fat embolism* and occur primarily on the upper body 2–3 days after a major injury. By using special fixatives, the emboli can be demonstrated in biopsy specimens of the petechiae. Rarely, emboli of tumor or thrombus are seen in patients with atrial myxomas and marantic endocarditis.

In the *Gardner-Diamond syndrome* (autoerythrocyte sensitivity), female patients develop large ecchymoses within areas of painful, warm erythema. Intradermal injections of autologous erythrocytes or phosphatidyl serine derived from the red cell membrane can reproduce the lesions in some patients; however, there are instances where a reaction is seen at an injection site of the forearm but not in the midback region. The latter has led some observers to view Gardner-Diamond syndrome as a cutaneous manifestation of severe emotional stress. More recently, the possibility of platelet dysfunction (as assessed via aggregation studies) has been raised. *Waldenström's hypergammaglobulinemic purpura* is a chronic disorder characterized by recurrent crops of petechiae and larger purpuric macules on the lower extremities. There are circulating complexes of IgG–anti-IgG molecules, and exacerbations are associated with prolonged standing or walking. Patients may have an underlying autoimmune connective tissue disease, e.g., Sjögren's syndrome.

Palpable purpura are further subdivided into vasculitic and embolic. In the group of vasculitic disorders, cutaneous small-vessel vasculitis, also known

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

respiratory infection. The majority of lesions are found on the lower extremities and buttocks. Systemic manifestations include fever, arthralgias (primarily of the knees and ankles), abdominal pain, gastrointestinal bleeding, and nephritis. Direct immunofluorescence examination shows deposits

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

Palpable purpura are further subdivided into vasculitic and embolic. In the group of vasculitic disorders, cutaneous small-vessel vasculitis, also known as *leukocytoclastic vasculitis* (LCV), is the one most commonly associated with palpable purpura **(Chap. 363)**. Underlying etiologies include drugs (e.g., antibiotics), infections (e.g., hepatitis C virus), and autoimmune connective tissue diseases (e.g., rheumatoid arthritis, Sjögren's syndrome, lupus). *Henoch-Schönlein purpura* (HSP) is a subtype of acute LCV that is seen more commonly in children and adolescents following an upper respiratory infection. The majority of lesions are found on the lower extremities and buttocks. Systemic manifestations include fever, arthralgias (primarily of the knees and ankles), abdominal pain, gastrointestinal bleeding, and nephritis. Direct immunofluorescence examination shows deposits of IgA within dermal blood vessel walls. Renal disease is of particular concern in adults with IgA vasculitis.

Several types of infectious emboli can give rise to palpable purpura. These embolic lesions are usually *irregular* in outline as opposed to the lesions of LCV, which are *circular* in outline. The irregular outline is indicative of a cutaneous infarct, and the size corresponds to the area of skin that received its blood supply from that particular arteriole or artery. The palpable purpura in LCV are circular because the erythrocytes simply diffuse out evenly from the postcapillary venules as a result of inflammation. Infectious emboli are most commonly due to gram-negative cocci (meningococcus, gonococcus), gram-negative rods (Enterobacteriaceae), and gram-positive cocci (*Staphylococcus*). Additional causes include *Rickettsia* and, in immunocompromised patients, *Aspergillus* and other opportunistic fungi.

The embolic lesions in *acute meningococcemia* are found primarily on the trunk, lower extremities, and sites of pressure, and a gunmetal-gray color often develops within them. Their size varies from a few millimeters to several centimeters, and the organisms can be cultured from the lesions. Associated findings include a preceding upper respiratory tract infection; fever; meningitis; DIC; and, in some patients, a deficiency of the terminal components of complement. In *disseminated gonococcal infection* (arthritis-dermatitis syndrome), a small number of inflammatory papules and vesicopustules, often with central purpura or hemorrhagic necrosis, are found on the distal extremities. Additional symptoms include arthralgias, tenosynovitis, and fever. To establish the diagnosis, a Gram stain of these lesions should be performed. *Rocky Mountain spotted fever* is a tick-borne disease that is caused by *Rickettsia rickettsii*. A several-day history of fever, chills, severe headache, and photophobia precedes the onset of the cutaneous eruption. The initial lesions are erythematous macules and papules on the wrists, ankles, palms, and soles. With time, the lesions spread centripetally and become purpuric.

Lesions of *ecthyma gangrenosum* begin as edematous, erythematous papules or plaques and then develop central purpura and necrosis. Bullae formation also occurs in these lesions, and they are frequently found in the girdle region. The organism that is classically associated with ecthyma gangrenosum is *Pseudomonas aeruginosa*, but other gram-negative rods such as *Klebsiella*, *Escherichia coli*, and *Serratia* can produce similar lesions. In immunocompromised hosts, the list of potential pathogens is expanded to include *Candida* and other opportunistic fungi (e.g., *Aspergillus*, *Fusarium*).

ULCERS

The approach to the patient with a cutaneous ulcer is outlined in **Table 58-17**. Peripheral vascular diseases of the extremities are reviewed in **Chap. 281**, as is Raynaud's phenomenon.

TABLE 58-17

Causes of Mucocutaneous Ulcers

I. Primary cutaneous disorders
   A. Peripheral vascular disease **(Chap. 281)**
      1. Venous
      2. Arterial[a]
   B. Livedoid vasculopathy in the setting of venous hypertension[b]
   C. Squamous cell carcinoma (e.g., within scars), basal cell carcinomas
   D. Infections, e.g., ecthyma caused by *Streptococcus* **(Chap. 148)**
   E. Physical, e.g., trauma, pressure
   F. Drugs, e.g., hydroxyurea
II. Systemic diseases
   A. Lower legs
      1. Small-vessel and medium-vessel vasculitis[c]
      4. Cholesterol emboli[a]
      5. Necrobiosis lipoidica[d]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

I. Primary cutaneous disorders
   A. Peripheral vascular disease **(Chap. 281)**
      1. Venous
      2. Arterial[a]
   B. Livedoid vasculopathy in the setting of venous hypertension[b]
   C. Squamous cell carcinoma (e.g., within scars), basal cell carcinomas
   D. Infections, e.g., ecthyma caused by *Streptococcus* **(Chap. 148)**
   E. Physical, e.g., trauma, pressure
   F. Drugs, e.g., hydroxyurea
II. Systemic diseases
   A. Lower legs
      1. Small-vessel and medium-vessel vasculitis[c]
      2. Hemoglobinopathies **(Chap. 98)**
      3. Cryoglobulinemia,[c] cryofibrinogenemia
      4. Cholesterol emboli[a,c]
      5. Necrobiosis lipoidica[d]
      6. Antiphospholipid syndrome **(Chap. 116)**
      7. Neuropathic[e] **(Chap. 403)**
      8. Panniculitis
      9. Kaposi's sarcoma, acral angiodermatitis (pseudo-Kaposi's sarcoma)
      10. Diffuse dermal angiomatosis
   B. Hands and feet
      1. Raynaud's phenomenon **(Chap. 281)**
      2. Buerger disease
   C. Generalized
      1. Pyoderma gangrenosum, but most commonly legs
      2. Calciphylaxis **(Chap. 410)**
      3. Infections, e.g., dimorphic fungi, leishmaniasis
      4. Lymphoma
   D. Face, especially perioral, and anogenital
      1. Chronic herpes simplex[f]
III. Mucosal
   A. Aphthae
   B. Drug-induced mucositis
   C. Behçet's disease **(Chap. 364)**
   D. Erythema multiforme major, Stevens-Johnson syndrome, TEN
   E. Primary blistering disorders **(Chap. 59)**
   F. Lupus erythematosus, lichen planus, lichenoid GVHD
   G. Inflammatory bowel disease
   H. Acute HIV infection
   I. Reactive arthritis

[a]Underlying atherosclerosis. [b]Also associated with underlying disorders that lead to hypercoagulability/thrombophilia, e.g., factor V Leiden, protein C dysfunction/deficiency, antiphospholipid antibodies. [c]Reviewed in section on purpura. [d]Reviewed in section on papulonodular skin lesions. [e]Favors plantar surface of the foot. [f]Sign of immunosuppression.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

lesions and livedo reticularis are found in association with *painful* ulcerations of the lower extremities. These ulcers are often slow to heal, but when they do, irregularly shaped white scars form. The majority of cases are secondary to venous hypertension, but possible underlying illnesses include

of the foot. [f]Sign of immunosuppression.

Abbreviations: GVHD, graft versus host disease; HIV, human immunodeficiency virus; TEN, toxic epidermal necrolysis.

Livedoid vasculopathy (livedoid vasculitis; atrophie blanche) represents a combination of a vasculopathy plus intravascular thrombosis. Purpuric lesions and livedo reticularis are found in association with *painful* ulcerations of the lower extremities. These ulcers are often slow to heal, but when they do, irregularly shaped white scars form. The majority of cases are secondary to venous hypertension, but possible underlying illnesses include disorders of hypercoagulability, for example, antiphospholipid syndrome and factor V Leiden **(Chaps. 117 and 357)**.

In *pyoderma gangrenosum*, the border of untreated active ulcers has a characteristic appearance consisting of an undermined necrotic violaceous edge and a peripheral erythematous halo. The ulcers often begin as pustules that then expand rather rapidly to a size as large as 20 cm. Although these lesions are most commonly found on the lower extremities, they can arise anywhere on the surface of the body, including at sites of trauma (pathergy). An estimated 30–50% of cases are idiopathic, and the most common associated disorders are ulcerative colitis and Crohn's disease. Less commonly, pyoderma gangrenosum is associated with seropositive rheumatoid arthritis, acute and chronic myelogenous leukemia, myelodysplasia, a monoclonal gammopathy (usually IgA), or an autoinflammatory disorder. Because the histology of pyoderma gangrenosum may be nonspecific (dermal infiltrate of neutrophils when in untreated state), the diagnosis requires clinicopathologic correlation, in particular, the exclusion of similar-appearing ulcers such as necrotizing vasculitis, Meleney's ulcer (synergistic infection at a site of trauma or surgery), dimorphic fungi, cutaneous amebiasis, spider bites, and factitial. In the myeloproliferative disorders, the ulcers may be more superficial with a pustulobullous border, and these lesions provide a connection between classic pyoderma gangrenosum and acute febrile neutrophilic dermatosis (Sweet syndrome).

FEVER AND RASH

The major considerations in a patient with a fever and a rash are inflammatory diseases versus infectious diseases. In the hospital setting, the most common scenario is a patient who has a drug rash plus a fever secondary to an underlying infection. However, it should be emphasized that a drug reaction can lead to both a cutaneous eruption and a fever ("drug fever"), especially in the setting of DRESS, AGEP, or serum sickness–like reaction. Additional inflammatory diseases that are often associated with a fever include pustular psoriasis, erythroderma, and Sweet syndrome. Lyme disease, secondary syphilis, and viral and bacterial exanthems (see "Exanthems," above) are examples of infectious diseases that produce a rash and a fever. Lastly, it is important to determine whether or not the cutaneous lesions represent septic emboli (see "Purpura," above). Such lesions usually have evidence of ischemia in the form of purpura, necrosis, or impending necrosis (gunmetal-gray color). In the patient with thrombocytopenia, however, purpura can be seen in inflammatory reactions such as morbilliform drug eruptions and infectious lesions.

FURTHER READING

Bolognia JL, Schaffer JV, Cerroni L (eds): *Dermatology*, 4th ed. Philadelphia, Elsevier, 2018.

Callen JP et al (eds): *Dermatological Signs of Systemic Disease*, 5th ed. Edinburgh, Elsevier, 2017.

Fazel N (ed): *Oral Signs of Systemic Disease*. Switzerland, Springer, 2019.

Kurtzman D: Rheumatologic dermatology. Clin Dermatol 36:439, 2018. [PubMed: 30047427]

Taylor SC et al (eds): *Taylor and Kelly's Dermatology for Skin of Color*, 2nd ed. New York, McGraw-Hill, 2016.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 58: Skin Manifestations of Internal Disease, Jean L. Bolognia; Jonathan S. Leventhal; Irwin M. Braverman
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 60: Cutaneous Drug Reactions

Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai

INTRODUCTION

Cutaneous reactions are the most frequent adverse reactions to medications, representing 10–15% of reported adverse drug reactions. Most are benign, but a few can be life threatening. Prompt recognition of severe reactions, drug withdrawal, and appropriate therapeutic interventions can minimize toxicity. This chapter focuses on adverse cutaneous reactions to systemic medications; it covers their incidence, patterns, and pathogenesis, and provides some practical guidelines on treatment, assessment of causality, and future use of drugs.

USE OF PRESCRIPTION DRUGS IN THE UNITED STATES

In the United States, more than 4 billion prescriptions for >60,000 drug products are dispensed annually. Hospital inpatients alone annually receive about 120 million courses of drug therapy, and half of adult Americans receive prescription drugs on a regular outpatient basis. Adverse effects of a prescription medication may result in 4.5 million urgent or emergency care visits and over 7000 deaths each year in the United States. Many patients use over-the-counter medicines that may cause adverse cutaneous reactions.

INCIDENCE OF CUTANEOUS REACTIONS

Several recent prospective studies reported that acute cutaneous reactions to drugs affect between 2.2 and 10 per 1000 hospitalized patients. Reactions usually occur a few days to 4 weeks after initiation of therapy.

In a series of 48,005 inpatients over a 20-year period, morbilliform rash (91%) and urticaria (6%) were the most frequent skin reactions, and antimicrobials, radiocontrast, and nonsteroidal anti-inflammatory drugs (NSAIDs) were the most common drug associations. Severe hypersensitivity reactions to medications have been reported to occur in between 1 in 1000 to 2 per million users, depending on the reaction type. Although rare, severe cutaneous reactions to drugs have an important impact on health because of significant sequelae; in addition, they may require hospitalization, increase the duration of hospital stay, or be life threatening. Some populations are at increased risk of drug reactions, including elderly patients, patients with autoimmune disease, hematopoietic stem cell transplant recipients, and those with acute Epstein-Barr virus (EBV) or human immunodeficiency virus (HIV) infection. The pathophysiology underlying this association is unknown but may be related to immune dysregulation. Individuals with advanced HIV disease (e.g., CD4 T lymphocyte count <200 cells/μL) have a 40- to 50-fold increased risk of adverse reactions to sulfamethoxazole **(Chap. 202)** and increased risk of severe hypersensitivity reactions.

In addition to acute eruptions, a variety of skin diseases can be induced or exacerbated by prolonged use of drugs (e.g., pruritus, pigmentation, nail or hair disorders, psoriasis, bullous pemphigoid, photosensitivity, and even cutaneous neoplasms). These drug reactions are not frequent; however, neither their incidence nor their impact on public health has been evaluated.

PATHOGENESIS OF DRUG REACTIONS

Adverse cutaneous responses to drugs can arise as a result of immunologic or nonimmunologic mechanisms.

NONIMMUNOLOGIC DRUG REACTIONS

Examples of nonimmunologic drug reactions are pigmentary changes due to dermal accumulation of medications or their metabolites, alteration of hair follicles by antimetabolites and signaling inhibitors, and lipodystrophy associated with metabolic effects of anti-HIV medications. These side effects are predictable and sometimes can be prevented.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

(e.g., urticaria, anaphylaxis), antibody-mediated reactions, immune complex deposition, and antigen-specific responses. Drug-specific CD4+ and CD8+

ACCESS Medicine®
Access Provided by: SILVERCHAIR

hair follicles by antimetabolites and signaling inhibitors, and lipodystrophy associated with metabolic effects of anti-HIV medications. These side effects are predictable and sometimes can be prevented.

IMMUNOLOGIC DRUG REACTIONS

Evidence suggests an immunologic basis for most acute drug eruptions. Drug reactions may result from immediate release of preformed mediators (e.g., urticaria, anaphylaxis), antibody-mediated reactions, immune complex deposition, and antigen-specific responses. Drug-specific CD4+ and CD8+ T cell clones can be derived from the blood or from skin lesions of patients with a variety of drug allergies, strongly suggesting that these T cells mediate drug allergy in an antigen-specific manner. Drug presentation to T cells is major histocompatibility complex (MHC)-restricted and likely involves drug-peptide complex recognition by specific T-cell receptors (TCRs).

Once a drug has induced an immune response, the phenotype of the reaction is determined by the nature of effectors: cytotoxic (CD8+) T cells in blistering and certain hypersensitivity reactions, chemokines for reactions mediated by neutrophils or eosinophils, and B cell collaboration for production of specific antibodies for urticarial reactions. Immunologic reactions have recently been classified into further subtypes that provide a useful framework for designating adverse drug reactions based on involvement of specific immune pathways **(Table 60-1)**.

TABLE 60-1

Classification of Adverse Drug Reactions Based on Immune Pathway

TYPE	KEY PATHWAY	KEY IMMUNE MEDIATORS	ADVERSE DRUG REACTION TYPE
Type I	IgE	IgE	Urticaria, angioedema, anaphylaxis
Type II	IgG-mediated cytotoxicity	IgG	Drug-induced hemolysis, thrombocytopenia (e.g., penicillin)
Type III	Immune complex	IgG + antigen	Vasculitis, serum sickness, drug-induced lupus
Type IVa	T lymphocyte–mediated macrophage inflammation	IFN-γ, TNF-α T_H1 cells	Tuberculin skin test, contact dermatitis
Type IVb	T lymphocyte–mediated eosinophil inflammation	IL-4, IL-5, IL-13 T_H2 cells Eosinophils	DIHS Morbilliform eruption
Type IVc	T lymphocyte–mediated cytotoxic T lymphocyte inflammation	Cytotoxic T lymphocytes Granzyme Perforin Granulysin (SJS/TEN] only)	SJS/TEN Morbilliform eruption
Type IVd	T lymphocyte–mediated neutrophil inflammation	CXCL8, IL-17, GM-CSF Neutrophils	AGEP

Abbreviations: AGEP, acute generalized exanthematous pustulosis; DIHS, drug-induced hypersensitivity syndrome; GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN, interferon; IL, interleukin; SJS, Stevens-Johnson syndrome; TEN, toxic epidermal necrolysis; TNF, tumor necrosis factor.

Immediate Reactions

Immediate reactions depend on the release of mediators of inflammation by tissue mast cells or circulating basophils. These mediators include

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

bronchospasm, laryngeal edema, and, occasionally, anaphylactic shock with hypotension and death. They occur within minutes of drug exposure. NSAIDs, including aspirin, and radiocontrast media are frequent causes of direct mast cell degranulation or anaphylactoid reactions, which can occur

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Immediate reactions depend on the release of mediators of inflammation by tissue mast cells or circulating basophils. These mediators include histamine, leukotrienes, prostaglandins, bradykinins, platelet-activating factor, enzymes, and proteoglycans. Drugs can trigger mediator release either directly ("anaphylactoid" reaction) or through IgE-specific antibodies. These reactions usually manifest in the skin and gastrointestinal, respiratory, and cardiovascular systems **(Chap. 353)**. Primary symptoms and signs include pruritus, urticaria, nausea, vomiting, abdominal cramps, bronchospasm, laryngeal edema, and, occasionally, anaphylactic shock with hypotension and death. They occur within minutes of drug exposure. NSAIDs, including aspirin, and radiocontrast media are frequent causes of direct mast cell degranulation or anaphylactoid reactions, which can occur on first exposure. Penicillins and muscle relaxants used in general anesthesia are the most frequent causes of IgE-dependent reactions to drugs, which require prior sensitization. Release of mediators is triggered when polyvalent drug protein conjugates cross-link IgE molecules fixed to sensitized cells. Certain routes of administration favor different clinical patterns (e.g., gastrointestinal effects from oral route, circulatory effects from intravenous route).

Immune Complex–Dependent Reactions

Serum sickness is produced by tissue deposition of circulating immune complexes with consumption of complement. It is characterized by fever, arthritis, nephritis, neuritis, edema, and an urticarial, papular, or purpuric rash **(Chap. 363)**. First described following administration of nonhuman sera, it currently occurs in the setting of monoclonal antibodies and similar medications. In classic serum sickness, symptoms develop 6 or more days after drug exposure, the latent period representing the time needed to synthesize antibody. Vasculitis, a relatively rare complication of drugs, may also be a result of immune complex deposition **(Chap. 363)**. Penicillin, cefaclor, amoxicillin, trimethoprim/sulfamethoxazole, and monoclonal antibodies such as infliximab, rituximab, and omalizumab may be associated with clinically similar "serum sickness–like" reactions (SSLR). The mechanism of this reaction is unknown but is unrelated to immune complex formation and complement activation, and systemic involvement is rare. Whereas serum sickness most commonly occurs in adults, SSLR is more frequently observed in children.

Delayed Hypersensitivity

While not completely understood, delayed hypersensitivity directed by drug-specific T cells is an important mechanism underlying the most common drug eruptions, that is, morbilliform eruptions, and also rare and severe forms such as drug-induced hypersensitivity syndrome (DIHS) (also known as drug rash with eosinophilia and systemic symptoms [DRESS]), acute generalized exanthematous pustulosis (AGEP), Stevens-Johnson syndrome (SJS), and toxic epidermal necrolysis (TEN) (Table 60-1). Drug-specific T cells have been detected in these types of drug eruptions. In TEN, skin lesions contain T lymphocytes reactive to autologous lymphocytes and keratinocytes in a drug-specific, human leukocyte antigen (HLA)-restricted, and perforin/granzyme-mediated pathway. In the case of carbamazepine, studies have identified cytotoxic T lymphocytes (CTLs) reactive to carbamazepine that use highly restricted V-alpha and V-beta TCR repertoires in patients with carbamazepine hypersensitivity that are not found in carbamazepine-tolerant individuals.

The mechanism(s) by which medications result in T cell activation is unknown. Two hypotheses prevail: first, that the antigens driving these reactions may be the native drug itself or components of the drug covalently complexed with endogenous proteins, presented in association with HLA molecules to T cells through the classic antigen presentation pathway or, alternatively, through direct interaction of the drug/metabolite with the TCR or peptide-loaded HLA (e.g., the pharmacologic interaction of drugs with immune receptors, or p-i hypothesis). Recent x-ray crystallography data characterizing binding between specific HLA molecules to drugs known to cause hypersensitivity reactions demonstrate unique alterations to the MHC peptide-binding groove, suggesting a molecular basis for T cell activation in the development of hypersensitivity reactions.

GENETIC FACTORS AND CUTANEOUS DRUG REACTIONS

 Genetic determinants may predispose individuals to severe drug reactions by affecting either drug metabolism or immune responses to drugs. Polymorphisms in cytochrome P450 enzymes, drug acetylation, methylation (such as thiopurine methyltransferase activity and azathioprine), and other forms of metabolism (such as glucose-6-phosphate dehydrogenase and dapsone) may increase susceptibility to drug toxicity or underdosing and increase risk for medication interactions, highlighting a role for differential pharmacokinetic or pharmacodynamic effects. The value of routine screening of P450 enzymes for prediction of cutaneous reactions has not been determined, though its cost-effectiveness in certain populations (e.g., patients with seizure disorder, depression) as well as patients considering specific therapies (e.g., tamoxifen, warfarin) has been suggested.

Associations between drug hypersensitivities and HLA haplotypes suggest a key role for immune mechanisms, especially those leading to skin involvement. Hypersensitivity to the anti-HIV medication abacavir is strongly associated with HLA-B*57:01 **(Chap. 202)**. In Taiwan, within a

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

association is strong, it is not sufficient to cause severe drug hypersensitivity reactions.

ACCESS Medicine

Access Provided by:
SILVERCHAIR

Associations between drug hypersensitivities and HLA haplotypes suggest a key role for immune mechanisms, especially those leading to skin involvement. Hypersensitivity to the anti-HIV medication abacavir is strongly associated with HLA-B*57:01 **(Chap. 202)**. In Taiwan, within a homogeneous Han Chinese population, a strong association was observed between SJS/TEN (but not DIHS) related to carbamazepine and HLA-B*15:02. In the same population, a strong association was found between HLA-B*58:01 and SJS, TEN, or DIHS related to allopurinol. These associations are drug and phenotype specific; that is, HLA-specific T cell stimulation by medications leads to distinct reactions. However, while this genetic association is strong, it is not sufficient to cause severe drug hypersensitivity reactions.

GLOBAL CONSIDERATIONS

Recognition of HLA associations with drug hypersensitivity has resulted in recommendations to screen high-risk populations. Genetic screening for HLA-B*57:01 to prevent abacavir hypersensitivity, which carries a 100% negative predictive value when patch test confirmed and 55% positive predictive value generalizable across races, is becoming the clinical standard of care worldwide (number needed to treat = 13). The U.S. Food and Drug Administration has recommended HLA-B*15:02 screening of Asian individuals prior to a new prescription of carbamazepine. The American College of Rheumatology has recommended HLA-B*58:01 screening of Han Chinese patients prescribed allopurinol. To date, screening for a single HLA (but not multiple HLA haplotypes) in specific populations has been determined to be cost-effective (e.g., HLA-B*1301 screening in Chinese patients with leprosy treated with dapsone). Genetic testing for specific HLA haplotypes and functional screening for TCR repertoire to identify patients at risk is becoming more widely available and heralds the era of personalized medicine and pharmacogenomics.

CLINICAL PRESENTATION OF CUTANEOUS DRUG REACTIONS

NONIMMUNE CUTANEOUS REACTIONS

Exacerbation or Induction of Dermatologic Diseases

A variety of drugs can exacerbate preexisting diseases or induce—or unmask—a disease that may or may not disappear after withdrawal of the inducing medication. For example, NSAIDs, lithium, beta blockers, tumor necrosis factor (TNF) antagonists, interferon (IFN) α, and angiotensin-converting enzyme (ACE) inhibitors can exacerbate plaque psoriasis, whereas antimalarials and withdrawal of systemic glucocorticoids can worsen pustular psoriasis. The situation of TNF-α inhibitors is unusual, as this class of medications is used to treat psoriasis; however, they may induce psoriasis (especially palmoplantar) in patients being treated for other conditions. Acne may be induced by glucocorticoids, androgens, lithium, and antidepressants. Follicular papular or pustular eruptions of the face and trunk resembling acne frequently occur with epidermal growth factor receptor (EGFR) antagonists, mitogen-activated protein kinase (MEK) inhibitors, and other targeted inhibitors. With EGFR antagonists, the severity of the eruption correlates with a better anticancer effect. This rash is typically responsive to and prevented by tetracycline antibiotics.

Several medications induce or exacerbate autoimmune disease. Checkpoint inhibitors induce a wide array of systemic autoimmune reactions, including in skin. Interleukin (IL) 2, IFN-α, and anti-TNF-α are associated with new-onset systemic lupus erythematosus (SLE). Drug-induced lupus is classically marked by antinuclear and antihistone antibodies and, in some cases, anti-double-stranded DNA (D-penicillamine, anti-TNF-α) or perinuclear antineutrophil cytoplasmic antibodies (p-ANCA) (minocycline). Subacute cutaneous lupus erythematosus (SCLE) can be induced by a growing list of drugs, including thiazide diuretics, proton pump inhibitors, TNF inhibitors, terbinafine, and minocycline. Drug-induced dermatomyositis may rarely occur with TNF inhibitors or capecitabine; hydroxyurea can induce skin findings of dermatomyositis. IFN and TNF inhibitors, as well as checkpoint inhibitors, can induce granulomatous disease and sarcoidosis. Autoimmune blistering diseases may be drug induced as well: pemphigus by D-penicillamine and ACE inhibitors; bullous pemphigoid by DPP4 inhibitors, furosemide, and PD-1 inhibitors; and linear IgA bullous dermatosis by vancomycin. Other medications may cause highly specific cutaneous reactions. Gadolinium contrast has been associated with nephrogenic systemic fibrosis, a condition of sclerosing skin with rare internal organ involvement; advanced renal compromise may be an important risk factor. Granulocyte colony-stimulating factor, azacitidine, all-*trans*-retinoic acid, the *FLT3* inhibitor class of drugs, and rarely levamisole-contaminated cocaine may induce neutrophilic dermatoses. In this setting, the hypothesis that a drug may be responsible should always be considered, even after the treatment is complete. In addition, reactions may develop in cases of long-term medication therapy due to changes in dosing or host metabolism. Resolution of the cutaneous reaction may be delayed upon discontinuation of the medication.

Photosensitivity Eruptions

Photosensitivity eruptions are usually most marked in sun-exposed areas, but they may extend to sun-protected areas. The mechanism is almost always phototoxicity. Phototoxic reactions resemble sunburn and can occur with first exposure to a drug. Blistering may occur in drug-related pseudoporphyria, most commonly with NSAIDs. The severity of the reaction depends on the tissue level of the drug, its efficiency as a photosensitizer

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

drugs less frequently implicated are chlorpromazine, thiazides, NSAIDs, and BRAF inhibitors. Voriconazole may result in severe photosensitivity, accelerated photoaging, and cutaneous carcinogenesis

Access Provided by:
SILVERCHAIR

always phototoxicity. Phototoxic reactions resemble sunburn and can occur with first exposure to a drug. Blistering may occur in drug-related pseudoporphyria, most commonly with NSAIDs. The severity of the reaction depends on the tissue level of the drug, its efficiency as a photosensitizer, and the extent of exposure to the activating wavelengths of ultraviolet (UV) light **(Chap. 61)**.

Common orally administered photosensitizing drugs include fluoroquinolones, tetracycline antibiotics, and trimethoprim/sulfamethoxazole. Other drugs less frequently implicated are chlorpromazine, thiazides, NSAIDs, and BRAF inhibitors. Voriconazole may result in severe photosensitivity, accelerated photoaging, and cutaneous carcinogenesis.

Because UV-A and visible light, which trigger these reactions, are not easily absorbed by nonopaque sunscreens and are transmitted through window glass, photosensitivity reactions may be difficult to block. Photosensitivity reactions abate with removal of either the drug or UV radiation, use of sunscreens that block UV-A light, and treatment of the reaction as one would a sunburn. Rarely, individuals develop persistent reactivity to light, necessitating long-term avoidance of sun exposure. Some chemotherapeutic agents, such as methotrexate, can induce a UV-recall reaction characterized by an erythematous, slightly scaly eruption at sites of prior severe sun exposure.

Pigmentation Changes

Drugs, either systemic or topical, may cause a variety of pigmentary changes in the skin by triggering melanocyte production of melanin (as in the case of oral contraceptives causing melasma) or due to deposition of drug or drug metabolites. Long-term minocycline and amiodarone may cause blue-gray pigmentation. Phenothiazine, gold, and bismuth result in gray-brown pigmentation of sun-exposed areas. Numerous cancer chemotherapeutic agents may be associated with characteristic patterns of pigmentation (e.g., bleomycin, busulfan, daunorubicin, cyclophosphamide, hydroxyurea, fluorouracil, and methotrexate). Clofazimine causes a drug-induced lipofuscinosis with characteristic red-brown coloration. Hyperpigmentation of the face, mucous membranes, and pretibial and subungual areas occurs with antimalarials. Quinacrine causes generalized yellow discoloration. Pigmentation changes may also occur in mucous membranes (busulfan, bismuth), conjunctiva (chlorpromazine, thioridazine, imipramine, clomipramine), nails (zidovudine, doxorubicin, cyclophosphamide, bleomycin, fluorouracil, hydroxyurea), hair, and teeth (tetracyclines).

Warfarin Necrosis of Skin

This rare reaction (0.01–0.1%) usually occurs between the third and tenth days of therapy with warfarin, usually in women. Common sites are breasts, thighs, and buttocks **(Fig. 60-1)**. Lesions are sharply demarcated, erythematous, or purpuric, and may progress to form large, hemorrhagic bullae with necrosis and eschar formation.

FIGURE 60-1

Warfarin necrosis involving the breasts. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Warfarin anticoagulation in protein C or S deficiency causes an additional reduction in already low circulating levels of endogenous anticoagulants, permitting hypercoagulability and thrombosis in the cutaneous microvasculature, with consequent areas of necrosis. Heparin-induced necrosis may have clinically similar features but is probably due to heparin-induced platelet aggregation with subsequent occlusion of blood vessels; it can affect

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Drug-Induced Hair Disorders

ACCESS Medicine®

Access Provided by:
SILVERCHAIR

permitting hypercoagulability and thrombosis in the cutaneous microvasculature, with consequent areas of necrosis. Heparin-induced necrosis may have clinically similar features but is probably due to heparin-induced platelet aggregation with subsequent occlusion of blood vessels; it can affect areas adjacent to the injection site or more distant sites if infused. Levamisole-tainted cocaine (and more recently, heroin) can induce similar skin necrosis; however, the distribution tends to involve the ears and cheeks predominantly, with stellate or retiform purpura. Patients may have abnormal white blood cell counts and may be dual P- and C-ANCA positive.

Drug-Induced Hair Disorders

DRUG-INDUCED HAIR LOSS

Medications may affect hair follicles at two different phases of their growth cycle: anagen (growth) or telogen (resting). *Anagen effluvium* occurs within days of drug administration, especially with antimetabolite or other chemotherapeutic drugs. In contrast, in *telogen effluvium*, the delay is 2–4 months following initiation of a new medication. Both present as diffuse, nonscarring alopecia most often reversible after discontinuation of the responsible agent.

A considerable number of drugs have been associated with hair loss. These include antineoplastic agents (alkylating agents, bleomycin, vinca alkaloids, platinum compounds), anticonvulsants (carbamazepine, valproate), beta blockers, antidepressants, antithyroid drugs, IFNs, oral contraceptives, and cholesterol-lowering agents.

DRUG-INDUCED HAIR GROWTH

Medications may also cause hair growth. Hirsutism is an excessive growth of terminal hair with masculine hair growth pattern in a female, most often on the face and trunk, due to androgenic stimulation of hormone-sensitive hair follicles (anabolic steroids, oral contraceptives, testosterone, corticotropin). Hypertrichosis is a distinct pattern of hair growth, not in a masculine pattern, typically located on the forehead and temporal regions of the face. Drugs responsible for hypertrichosis include anti-inflammatory drugs, glucocorticoids, vasodilators (diazoxide, minoxidil), diuretics (acetazolamide), anticonvulsants (phenytoin), immunosuppressive agents (cyclosporine A), psoralens, and zidovudine.

Changes in hair color or structure are uncommon adverse effects from medications. Hair discoloration may occur with chloroquine, IFN-α, chemotherapeutic agents, and tyrosine kinase inhibitors. Changes in hair structure have been observed in patients given EGFR inhibitors, BRAF inhibitors, tyrosine kinase inhibitors, and acitretin.

Drug-Induced Nail Disorders

Drug-related nail disorders usually involve all 20 nails and need months to resolve after withdrawal of the medication. The pathogenesis is most often toxic. Drug-induced nail changes include Beau's line (transverse depression of the nail plate), onycholysis (detachment of the distal part of the nail plate), onychomadesis (detachment of the proximal part of the nail plate), pigmentation, and paronychia (inflammation of periungual skin).

ONYCHOLYSIS

Onycholysis occurs with tetracyclines, fluoroquinolones, retinoids, NSAIDs, and others, including many chemotherapeutic agents, and may be triggered by exposure to sunlight.

ONYCHOMADESIS

Onychomadesis is caused by temporary arrest of nail matrix mitotic activity. Common drugs reported to induce onychomadesis include carbamazepine, lithium, retinoids, and chemotherapeutic agents such as taxanes.

PARONYCHIA

Paronychia and multiple pyogenic granulomas with progressive and painful periungual abscess of fingers and toes are side effects of systemic retinoids, lamivudine, indinavir, and anti-EGFR monoclonal antibodies.

NAIL DISCOLORATION

Some drugs—including anthracyclines, taxanes, fluorouracil, psoralens, and zidovudine—may induce nail bed hyperpigmentation through melanocyte stimulation. It appears to be reversible and dose dependent.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Because many agents used in cancer chemotherapy inhibit cell division, rapidly proliferating elements of the skin, including hair, mucous membranes, and appendages, are sensitive to their effects. A broad spectrum of chemotherapy-related skin toxicities has been reported, including neutrophilic

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Some drugs—including anthracyclines, taxanes, fluorouracil, psoralens, and zidovudine—may induce nail bed hyperpigmentation through melanocyte stimulation. It appears to be reversible and dose dependent.

Toxic Erythema of Chemotherapy and Other Chemotherapy Reactions

Because many agents used in cancer chemotherapy inhibit cell division, rapidly proliferating elements of the skin, including hair, mucous membranes, and appendages, are sensitive to their effects. A broad spectrum of chemotherapy-related skin toxicities has been reported, including neutrophilic eccrine hidradenitis, sterile cellulitis, exfoliative dermatitis, and flexural erythema; recent nomenclature classifies these under the unifying diagnosis of toxic erythema of chemotherapy (TEC) **(Fig. 60-2).** Acral erythema is marked by dysesthesia and an erythematous, edematous eruption of the palms and soles. Common causes include cytarabine, doxorubicin, methotrexate, hydroxyurea, fluorouracil, and capecitabine.

FIGURE 60-2

Toxic erythema of chemotherapy. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The recent introduction of many new monoclonal antibody and small molecular signaling inhibitors for the treatment of cancer has been accompanied by numerous reports of skin and hair toxicity; only the most common of these are mentioned here. EGFR antagonists induce follicular eruptions and nail toxicity after a mean interval of 10 days in a majority of patients. Xerosis, eczematous eruptions, acneiform eruptions, and pruritus are common. Erlotinib is associated with marked hair textural changes. Sorafenib, a tyrosine kinase inhibitor, may result in follicular eruptions and focal bullous eruptions at palmoplantar, flexural sites or areas of frictional pressure. BRAF inhibitors are associated with photosensitivity, palmoplantar hyperkeratosis, hair curling, dyskeratotic (Grover's-like) rash, hyperkeratotic benign cutaneous neoplasms, and keratoacanthoma-like squamous cell carcinomas. Rash, pruritus, and vitiliginous depigmentation have been reported in association with ipilimumab (anti-CTLA4) treatment. Up to 50% of patients experience immune-mediated skin eruptions, including granulomatous reactions, dermatomyositis, panniculitis, and vasculitis. The checkpoint inhibitor class of drugs (including anti-CTLA4, anti-PD-1, and anti-PD-L1 agents) can induce a wide range of cutaneous eruptions beyond vitiligo, including lichenoid, eczematous, granulomatous, papulosquamous, and panniculitis eruptions.

IMMUNE CUTANEOUS REACTIONS: COMMON

Maculopapular Eruptions

Morbilliform or maculopapular eruptions **(Fig. 60-3)** are the most common of all drug-induced reactions, often start on the trunk or intertriginous areas, and consist of blanching erythematous macules and papules that are symmetric and confluent. Nonblanching, dusky, or bright-red macules as well as mucosal involvement should raise concern for a more severe reaction. Facial involvement in morbilliform eruptions is also uncommon, and the presence of extensive facial lesions with facial edema suggests DIHS. Diagnosis of morbilliform eruptions is rarely assisted by laboratory testing or skin

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Morbilliform drug eruption. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine,*

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

well as mucosal involvement should raise concern for a more severe reaction. Facial involvement in morbilliform eruptions is also uncommon, and the presence of extensive facial lesions with facial edema suggests DIHS. Diagnosis of morbilliform eruptions is rarely assisted by laboratory testing or skin biopsy.

FIGURE 60-3

Morbilliform drug eruption. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Morbilliform eruptions may be associated with moderate to severe pruritus and fever. A viral exanthem is another differential diagnostic consideration, especially in children, and graft-versus-host disease is also a consideration in the proper clinical setting. Absence of enanthems; absence of ear, nose, throat, and upper respiratory tract symptoms; and polymorphism of the skin lesions support a drug rather than a viral eruption. Common offenders include aminopenicillins, cephalosporins, antibacterial sulfonamides, allopurinol, and antiepileptic drugs. Beta blockers, calcium channel blockers, and ACE inhibitors are rarely the culprit; however, any drug can cause a morbilliform exanthem. Certain medications carry very high rates of morbilliform eruption, including nevirapine and lamotrigine, even in the absence of DIHS reactions. Lamotrigine morbilliform rash is associated with higher starting doses, rapid dose escalation, concomitant use of valproate (which increases lamotrigine levels and half-life), and use in children.

Maculopapular reactions usually develop within 1 week of initiation of therapy and last less than 2 weeks. Occasionally, these eruptions resolve despite continued use of the responsible drug. Because the eruption may also worsen, the suspect drug should be discontinued unless it is essential. It is important to note that the rash may continue to progress for a few days up to 1 week following medication discontinuation. Oral antihistamines and emollients may help relieve pruritus. Short courses of potent topical glucocorticoids can reduce inflammation and symptoms. Systemic glucocorticoid treatment is rarely indicated.

Pruritus

Pruritus is associated with almost all drug eruptions and, in some cases, may represent the only symptom of the adverse cutaneous reaction. It may be alleviated by antihistamines such as hydroxyzine or diphenhydramine. Pruritus stemming from specific medications may require distinct treatment, such as selective opiate antagonists for opiate-related pruritus.

Urticaria/Angioedema/Anaphylaxis

Urticaria, the second most frequent type of cutaneous reaction to drugs, is characterized by pruritic, red wheals of varying size rarely lasting more than 24 hours. It has been observed in association with nearly all drugs, most frequently ACE inhibitors, aspirin, NSAIDs, penicillin, and blood products. However, medications account for no more than 10–20% of acute urticaria cases. Deep edema within dermal and subcutaneous tissues is known as angioedema and may involve respiratory and gastrointestinal mucous membranes. Urticaria and angioedema may be part of a life-threatening anaphylactic reaction.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

within minutes. Immune complex–induced urticaria associated with serum sickness reactions usually occurs 6–12 days after first exposure. In this syndrome, the urticarial eruption (typically polycyclic plaques over distal joints) may be accompanied by fever, hematuria, arthralgias, hepatic

Access Provided by:
SILVERCHAIR

angioedema and may involve respiratory and gastrointestinal mucous membranes. Urticaria and angioedema may be part of a life-threatening anaphylactic reaction.

Drug-induced urticaria may be caused by three mechanisms: an IgE-dependent mechanism, circulating immune complexes (serum sickness), and nonimmunologic activation of effector pathways. IgE-dependent urticarial reactions usually occur within 36 hours of drug exposure but can occur within minutes. Immune complex–induced urticaria associated with serum sickness reactions usually occurs 6–12 days after first exposure. In this syndrome, the urticarial eruption (typically polycyclic plaques over distal joints) may be accompanied by fever, hematuria, arthralgias, hepatic dysfunction, and neurologic symptoms. Certain drugs, such as NSAIDs, ACE inhibitors, angiotensin II antagonists, radiographic dye, and opiates, may induce urticarial reactions, angioedema, and anaphylaxis in the absence of drug-specific antibodies through direct mast-cell degranulation.

Radiocontrast agents are a common cause of urticaria and, in rare cases, can cause anaphylaxis. High-osmolality radiocontrast media are about five times more likely to induce urticaria (1%) or anaphylaxis than are newer low-osmolality media. About one-third of those with mild reactions to previous exposure react on reexposure. Pretreatment with prednisone and diphenhydramine reduces reaction rates.

The treatment of urticaria or angioedema depends on the severity of the reaction. In severe cases with respiratory or cardiovascular compromise, epinephrine and intravenous glucocorticoids are the mainstay of therapy. For patients with urticaria without symptoms of angioedema or anaphylaxis, drug withdrawal and oral antihistamines are usually sufficient. Future drug avoidance is recommended; rechallenge, especially in individuals with severe reactions, should only occur in an intensive care setting.

Anaphylactoid Reactions

Vancomycin is associated with red man syndrome, a histamine-related anaphylactoid reaction characterized by flushing, diffuse maculopapular eruption, and hypotension. In rare cases, cardiac arrest may be associated with rapid intravenous (IV) infusion of the medication.

Irritant/Allergic Contact Dermatitis

Patients using topical medications may develop an irritant or allergic contact dermatitis to the medication itself or to a preservative or other component of the formulation. Reactions to neomycin sulfate, bacitracin, and polymyxin B are common. Contact dermatitis may be seen to adhesive tapes, leading to irritation or blisters around ports and IV sites **(Fig. 60-4)**. Harsh disinfectant skin cleansers may lead to localized irritant dermatitis.

FIGURE 60-4

Allergic contact dermatitis (bullous) due to adhesive tape. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

violaceous erythema and central bulla **(Fig. 60-5)**. Hyperpigmentation often results after resolution of the acute inflammation. With rechallenge, the process recurs in the same (fixed) location but may spread to new areas as well. Lesions often involve the lips, hands, legs, face, genitalia, and oral

Copyright © McGraw Hill. All rights reserved.

Access Provided by:
SILVERCHAIR

Fixed Drug Eruptions

These less common reactions are characterized by one or more sharply demarcated, dull red to brown lesions, sometimes with central dusky violaceous erythema and central bulla **(Fig. 60-5)**. Hyperpigmentation often results after resolution of the acute inflammation. With rechallenge, the process recurs in the same (fixed) location but may spread to new areas as well. Lesions often involve the lips, hands, legs, face, genitalia, and oral mucosa, and cause a burning sensation. Most patients have multiple lesions. Fixed drug eruptions have been associated with pseudoephedrine (frequently a nonpigmenting reaction), phenolphthalein (in laxatives), sulfonamides, tetracyclines, NSAIDs, barbiturates, and others.

FIGURE 60-5

Fixed drug eruption. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

IMMUNE CUTANEOUS REACTIONS: RARE AND SEVERE

Drug-Induced Hypersensitivity Syndrome

DIHS is a systemic drug reaction also known as DRESS (drug reaction with eosinophilia and systemic symptoms) syndrome; because eosinophilia is not always present, the term *DIHS* is preferred. Clinically, DIHS presents with a prodrome of fever and flu-like symptoms for several days, followed by the appearance of a diffuse morbilliform eruption, usually involving the face **(Fig. 60-6)**. Facial swelling and hand/foot swelling are often present. Systemic manifestations include lymphadenopathy, fever, and leukocytosis (often with eosinophilia or atypical lymphocytosis), as well as hepatitis, nephritis, pneumonitis, myositis, and gastroenteritis, in descending order. Distinct patterns of timing of onset and organ involvement may exist. For example, allopurinol classically induces DIHS with renal involvement; cardiac and lung involvement are more common with minocycline; gastrointestinal involvement is almost exclusively seen with abacavir; and some medications typically do not induce eosinophilia (abacavir, dapsone, lamotrigine). The cutaneous reaction usually begins 2–8 weeks after the drug is started and persists after drug cessation. Signs and symptoms may continue for several weeks, especially those associated with hepatitis. The eruption recurs with rechallenge, and cross-reactions among aromatic anticonvulsants, including phenytoin, carbamazepine, and phenobarbital, are common. Other drugs causing DIHS include antibacterial sulfonamides and other antibiotics. Hypersensitivity to reactive drug metabolites, hydroxylamine for sulfamethoxazole and arene oxide for aromatic anticonvulsants, may be involved in the pathogenesis of DIHS. Recent research suggests that inciting drugs may reactivate quiescent human herpes viruses, including herpesviruses 6 and 7, EBV, and cytomegalovirus (CMV), resulting in expansion of viral-specific CD8+ T lymphocytes and subsequent end-organ damage. Viral reactivation may be associated with a worse clinical prognosis. Mortality rates as high as 10% have been reported, with most fatalities resulting from liver failure. Systemic glucocorticoids (1.5–2 mg/kg/d prednisone equivalent) should be started and tapered slowly over 8–12

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

complications of myocarditis, patients should undergo cardiac evaluation in cases of severe DIHS or if heart involvement is suspected due to

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

anticonvulsants, including phenytoin, carbamazepine, and phenobarbital, are common. Other drugs causing DIHS include antibacterial sulfonamides

end-organ damage. Viral reactivation may be associated with a worse clinical prognosis. Mortality rates as high as 10% have been reported, with most fatalities resulting from liver failure. Systemic glucocorticoids (1.5–2 mg/kg/d prednisone equivalent) should be started and tapered slowly over 8–12 weeks, during which time clinical symptoms and labs (including complete blood count with differential, basic metabolic panel, and liver function tests) should be followed carefully. A steroid-sparing agent such as mycophenolate mofetil, IV immunoglobulin, or cyclosporine may be indicated in cases of rapid recurrence upon steroid taper. In all cases, immediate withdrawal of the suspected culprit drug is required. Given the severe long-term complications of myocarditis, patients should undergo cardiac evaluation in cases of severe DIHS or if heart involvement is suspected due to hypotension or arrhythmia. Patients should be closely monitored for resolution of organ dysfunction and for development of late-onset autoimmune thyroiditis and diabetes (up to 6 months).

FIGURE 60-6

Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS). *(Courtesy of Gildo Micheletti, MD.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis

SJS and *TEN* are characterized by blisters and mucosal/epidermal detachment resulting from full-thickness epidermal necrosis in the absence of substantial dermal inflammation. The term *Stevens-Johnson syndrome* (SJS) describes cases in which the total body surface area of blistering and eventual detachment is <10% **(Fig. 60-7)**. The term *Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) overlap* is used to describe cases with 10–30% epidermal detachment **(Fig. 60-8)**, and the term *toxic epidermal necrolysis (TEN)* is used to describe cases with >30% detachment **(Figs. 60-9 and 60-10)**.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Stevens-Johnson syndrome (SJS). *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

cases with 10–30% epidermal detachment **(Fig. 60-8)**, and the term *toxic epidermal necrolysis (TEN)* is used to describe cases with >30% detachment **(Figs. 60-9 and 60-10)**.

FIGURE 60-7

Stevens-Johnson syndrome (SJS). *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 60-8

SJS-TEN overlap. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 60-9

Toxic epidermal necrolysis, hand. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 60-10

Toxic epidermal necrolysis. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Copyright © McGraw Hill. All rights reserved.

Other blistering eruptions with concomitant mucositis may be confused with SJS/TEN. Erythema multiforme (EM) associated with herpes simplex virus



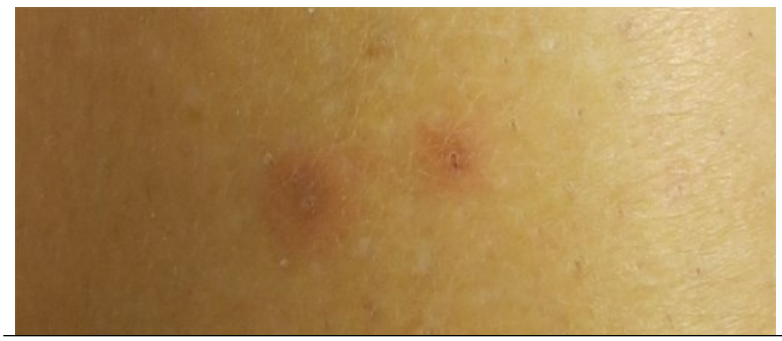
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Other blistering eruptions with concomitant mucositis may be confused with SJS/TEN. Erythema multiforme (EM) associated with herpes simplex virus is characterized by painful mucosal erosions and target lesions, typically with an acral distribution and limited skin detachment. *Mycoplasma* and other respiratory infections in children cause a clinically distinct presentation with prominent mucositis and limited cutaneous involvement. The term *reactive infectious mucocutaneous eruption (RIME)* has been proposed to help differentiate this clinical entity, which some believe may be the syndrome originally described by Stevens and Johnson.

Patients with SJS/TEN initially present with fever >39°C (102.2°F); sore throat; conjunctivitis; and acute onset of painful dusky, atypical, target-like lesions **(Fig. 60-11)**. Intestinal and upper respiratory tract involvement are associated with a poor prognosis, as are older age and greater extent of epidermal detachment. At least 10% of those with SJS and 30% of those with TEN die from the disease. Drugs that most commonly cause SJS/TEN are sulfonamides, allopurinol, antiepileptics (e.g., lamotrigine, phenytoin, carbamazepine), oxicam NSAIDs, β-lactam and other antibiotics, and nevirapine. Frozen-section skin biopsy may aid in rapid diagnosis.

FIGURE 60-11

Target-like lesion in SJS. *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

At this time, there is no consensus on the most effective treatment for SJS/TEN. The best outcomes stem from early diagnosis, immediate discontinuation of the suspected drug, and meticulous supportive therapy in an intensive care or burn unit. Fluid management, atraumatic wound care, infection prevention and treatment, and ophthalmologic and respiratory support are critical. Early administration of systemic glucocorticoids, intravenous immunoglobulin, cyclosporine, or etanercept may improve disease outcomes, but randomized studies to evaluate potential therapies are lacking and difficult to perform.

Pustular Eruptions

AGEP is a rare reaction pattern affecting 3–5 people per million per year. It is thought to be secondary to medication exposure in >90% of cases **(Fig. 60-12)**. Patients typically present with diffuse erythema or erythroderma, as well as high spiking fevers and leukocytosis with neutrophilia. One to two days later, innumerable pinpoint pustules develop overlying the erythema. The pustules are most pronounced in body fold areas; however, they may become generalized and, when coalescent, can lead to superficial erosion. In such cases, differentiating the eruption from SJS in its initial stages may be difficult, although in AGEP, any erosions tend to be more superficial, and prominent mucosal involvement is lacking. Skin biopsy shows collections of neutrophils and sparse necrotic keratinocytes in the upper part of the epidermis, unlike the full-thickness epidermal necrosis that characterizes SJS. Before the pustules appear, AGEP may also mimic DIHS due to the prominent fever and erythroderma.

FIGURE 60-12

Acute generalized exanthematous pustulosis.



Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The principal differential diagnosis for AGEP is acute pustular psoriasis, which has an identical clinical and histologic appearance. Many patients with AGEP have a personal or family history of psoriasis. AGEP classically begins within 24–48 hours of drug exposure, although it may occur as much as 1–2 weeks later. β-Lactam antibiotics, calcium channel blockers, macrolide antibiotics, and other inciting agents (including radiocontrast and dialysates) have been reported. Patch testing with the responsible drug often results in a localized pustular eruption.

Overlap Hypersensitivity Syndromes

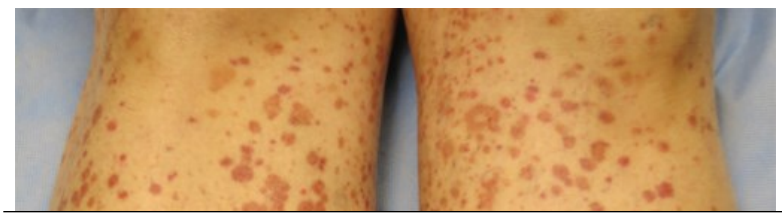
An important concept in the clinical approach to severe drug eruptions is the presence of "overlap syndromes," most notably DIHS with TEN-like features, DIHS with pustular eruption (AGEP-like), and AGEP with TEN-like features. In several case series of AGEP, 50% of cases had TEN-like or DRESS-like features, and 20% of cases had mucosal involvement resembling SJS/TEN. In one study, up to 20% of all severe drug eruptions had overlap features, suggesting that AGEP, DIHS, and SJS/TEN represent a clinical spectrum with some common pathophysiologic mechanisms. Designation of a single diagnosis based on cutaneous and extracutaneous involvement may not always be possible in cases of hypersensitivity; in such instances, treatment should be geared toward addressing the dominant clinical features. The timing of rash onset with respect to drug administration, which is usually much more delayed in DIHS, and the presence of systemic manifestations such as hepatitis are helpful clues to that diagnosis.

Vasculitis

Cutaneous small-vessel vasculitis (CSVV) typically presents with purpuric papules and macules involving the lower extremities and other dependent areas **(Fig. 60-13) (Chap. 363)**. Pustular and hemorrhagic vesicles as well as rounded ulcers also occur. Importantly, vasculitis may involve other organs, including the kidneys, joints, gastrointestinal tract, and lungs, necessitating a thorough clinical evaluation for systemic involvement. Drugs are implicated as a cause of roughly 15% of all cases of small-vessel vasculitis. Antibiotics, particularly β-lactams, are commonly implicated; however, almost any drug can cause vasculitis. Vasculitis may also be idiopathic or due to underlying infection, connective tissue disease, or (rarely) malignancy.

FIGURE 60-13

Cutaneous small-vessel vasculitis (CSVV, leukocytoclastic vasculitis). *(Source: JL Jameson, AS Fauci, DL Kasper, SL Hauser, DL Longo, J Loscalzo: Harrison's Principles of Internal Medicine, 20th Edition Copyright © McGraw Hill Education. All rights reserved.)*



Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Rare but important types of drug-induced vasculitis include drug-induced ANCA vasculitis. Such patients commonly present with cutaneous manifestations but can develop the full range of symptoms associated with ANCA-associated vasculitis, including crescentic glomerulonephritis and alveolar hemorrhage. Propylthiouracil, methimazole, and hydralazine are common culprits. Drug-induced polyarteritis nodosa has been associated with long-term exposure to minocycline. The presence of perivascular eosinophils on skin biopsy can be a clue to possible drug etiology.

MANAGEMENT OF THE PATIENT WITH SUSPECTED DRUG ERUPTION

There are four main questions to answer regarding a suspected drug eruption:

1. Is the observed rash caused by a medication?
2. Is the reaction severe or evolving with systemic involvement?
3. Which drug or drugs are suspected, and should they be withdrawn?
4. What recommendation can be made for future medication use?

EARLY DIAGNOSIS OF SEVERE ERUPTIONS

Rapid recognition of potentially serious or life-threatening reactions is paramount. In this regard, a suspected drug eruption is best defined initially by what it is not (e.g., SJS/TEN, DIHS). **Table 60-2** lists clinical and laboratory features that, if present, suggest the presence of a severe reaction. **Table**

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

what it is not (e.g., SJS/TEN, DIHS). **Table 60-2** lists clinical and laboratory features that, if present, suggest the presence of a severe reaction. **Table 60-3** lists the most important of these reactions, along with their key features and commonly associated medications. Any concern for a serious reaction should prompt immediate consultation with a dermatologist and/or referral of the patient to a specialized center.

TABLE 60-2

Clinical and Laboratory Findings Suggestive of Severe Cutaneous Adverse Drug Reaction

Cutaneous
Generalized erythema Facial edema Skin pain Palpable purpura Dusky or target-like lesions Skin necrosis Blisters or epidermal detachment Positive Nikolsky sign Mucous membrane erosions Swelling of lips or tongue
General
High fever Enlarged lymph nodes Arthralgias or arthritis Shortness of breath, hoarseness, wheezing, hypotension
Laboratory Results
Eosinophil count >1000/μL Lymphocytosis with atypical lymphocytes Abnormal liver or kidney function tests

Source: From The New England Journal of Medicine. Jean Claude Roujeau, Robert S. Stern: Severe Adverse Cutaneous Reactions to Drugs 331: 1272. Copyright © 1994 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

TABLE 60-3

Clinical Features of Severe Cutaneous Drug Reactions

DIAGNOSIS	MUCOSAL LESIONS	TYPICAL SKIN LESIONS	FREQUENT SIGNS AND SYMPTOMS	MOST COMMON CULPRIT DRUGS
Stevens-Johnson syndrome (SJS)	Erosions usually at two or more sites	Small blisters form from dusky macules or atypical targets; rare areas of confluence; detachment ≤10% body surface area	Most cases involve fever	Sulfonamides, anticonvulsants, allopurinol, nonsteroidal anti-inflammatory drugs (NSAIDs)
Toxic epidermal necrolysis (TEN)[a]	Erosions usually at	Individual lesions like those seen in SJS; confluent dusky erythema; large sheets of	Nearly all cases involve fever, “acute skin failure,”	Same as for SJS
Drug-induced hypersensitivity syndrome/drug rash with	Mucositis reported	Diffuse, deep red morbilliform eruption with facial involvement; facial and acral	Fever, lymphadenopathy, hepatitis, nephritis,	Anticonvulsants, sulfonamides, allopurinol,

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

1994 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

TABLE 60-3

Clinical Features of Severe Cutaneous Drug Reactions

DIAGNOSIS	MUCOSAL LESIONS	TYPICAL SKIN LESIONS	FREQUENT SIGNS AND SYMPTOMS	MOST COMMON CULPRIT DRUGS
Stevens-Johnson syndrome (SJS)	Erosions usually at two or more sites	Small blisters form from dusky macules or atypical targets; rare areas of confluence; detachment ≤10% body surface area	Most cases involve fever	Sulfonamides, anticonvulsants, allopurinol, nonsteroidal anti-inflammatory drugs (NSAIDs)
Toxic epidermal necrolysis (TEN)[a]	Erosions usually at two or more sites	Individual lesions like those seen in SJS; confluent dusky erythema; large sheets of necrotic epidermis; total detachment of >30% body surface area	Nearly all cases involve fever, "acute skin failure," leukopenia	Same as for SJS
Drug-induced hypersensitivity syndrome/drug rash with eosinophilia and systemic symptoms (DIHS/DRESS)	Mucositis reported in as many as 30%	Diffuse, deep red morbilliform eruption with facial involvement; facial and acral swelling	Fever, lymphadenopathy, hepatitis, nephritis, myocarditis, eosinophilia, atypical lymphocytosis	Anticonvulsants, sulfonamides, allopurinol, minocycline
Acute generalized exanthematous pustulosis (AGEP)	Oral erosions in perhaps 20%	Innumerable pinpoint pustules overlying a diffuse erythematous eruption; may develop superficial erosions	High fever, leukocytosis (neutrophilia), hypocalcemia	β-Lactam antibiotics, calcium channel blockers, macrolide antibiotics
Serum sickness or serum sickness–like reaction	Absent	Urticarial serpiginous or polycyclic rash; purpuric eruption along the sides of the feet and hands is characteristic	Fever, arthralgias	Antithymocyte globulin, cephalosporins, monoclonal antibodies
Anticoagulant-induced necrosis	Infrequent	Purpura and necrosis, especially of central, fatty areas	Pain in affected areas	Warfarin, heparin
Angioedema	Often involved	Urticaria or swelling of the central face, other areas	Respiratory distress, cardiovascular collapse	Angiotensin-converting enzyme (ACE) inhibitors, NSAIDs, contrast dye

[a]Overlap of SJS and TEN have features of both, and attachment of 10–30% of body surface area may occur.

Source: From The New England Journal of Medicine. Jean Claude Roujeau, Robert S. Stern: Severe Adverse Cutaneous Reactions to Drugs 331: 1272. Copyright © 1994 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.

CONFIRMATION OF DRUG REACTION

The probability of drug etiology varies with the pattern of the reaction. Only fixed drug eruptions are always drug-induced. Morbilliform eruptions are usually viral in children and drug-induced in adults. Among severe reactions, drugs account for 10–20% of anaphylaxis and vasculitis and between 70% and 90% of AGEP, DIHS, SJS, and TEN. Skin biopsy helps characterize the reaction but does not indicate drug causality. Blood counts and liver and renal function tests are important for evaluating organ involvement. The association of mild elevation of liver enzymes and high eosinophil count is

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

WHAT DRUG(S) TO SUSPECT AND WITHDRAW

renal function tests are important for evaluating organ involvement. The association of mild elevation of liver enzymes and high eosinophil count is frequent but not specific for a drug reaction. Blood tests that could identify an alternative cause, serologic tests (to rule out drug-induced lupus), and serology or polymerase chain reaction for infections may be of great importance to determine a cause.

WHAT DRUG(S) TO SUSPECT AND WITHDRAW

Most cases of drug eruptions occur during the first course of treatment with a new medication. A notable exception is IgE-mediated urticaria and anaphylaxis that need presensitization and develop a few minutes to a few hours after rechallenge. Characteristic timing of onset following drug administration is as follows: 4–14 days for morbilliform eruption, 2–4 days for AGEP, 5–28 days for SJS/TEN, and 14–48 days for DIHS. A drug chart, compiling information of all current and past medications/supplements and the timing of administration relative to the rash, is a key diagnostic tool for identifying the inciting drug. Medications introduced for the first time in the relevant time frame are prime suspects. Two other important elements to suspect causality at this stage are (1) previous experience with the drug (or related members of the same pharmacologic class) and (2) alternative etiologic candidates.

The decision to continue or discontinue any medication depends on the severity of the reaction, the severity of the primary disease undergoing treatment, the degree of suspicion of causality, and the feasibility of finding an alternative safer treatment. In any potentially fatal drug reaction, elimination of all possible suspect drugs or unnecessary medications should be immediately attempted. Some rashes may resolve when “treating through” a benign drug-related eruption. The decision to treat through an eruption should, however, remain the exception and withdrawal of every suspect drug the general rule. On the other hand, drugs that are not suspected and are important for the patient (e.g., antihypertensive agents) generally should not be quickly withdrawn. This approach may permit judicious use of these agents in the future.

RECOMMENDATION FOR FUTURE USE OF DRUGS

The aims are to (1) prevent the recurrence of the drug eruption and (2) avoid compromising future treatment by inaccurately excluding otherwise useful medications.

A thorough assessment of drug causality is based on timing of the reaction, evaluation of other possible causes, and effect of drug withdrawal or continuation. The RegiSCAR group has proposed the Algorithm of Drug Causality for Epidermal Necrolysis (ALDEN) to rank likelihood of drug causality in SJS/TEN; validation of this and other instruments, such as the Naranjo adverse drug reaction probability scale, is limited. Medication(s) with a “definite” or “probable” causality should be contraindicated, a warning card or medical alert tag (e.g., wristband) should be given to the patient, and the drugs should be listed in the patient’s medical chart as allergies.

CROSS-SENSITIVITY

Because of possible cross-sensitivity among chemically related drugs, many physicians recommend avoidance of not only the medication that induced the reaction but also all drugs of the same pharmacologic class.

There are two types of cross-sensitivity. Reactions that depend on a pharmacologic interaction may occur with all drugs that target the same pathway, whether the drugs are structurally similar or not. This is the case with angioedema caused by NSAIDs and ACE inhibitors. In this situation, the risk of recurrence varies from drug to drug in a particular class; however, avoidance of all drugs in the class is usually recommended. Immune recognition of structurally related drugs is the second mechanism by which cross-sensitivity occurs. A classic example is hypersensitivity to aromatic antiepileptics (barbiturates, phenytoin, carbamazepine) with up to 50% reaction to a second drug in patients who reacted to one. For other drugs, in vitro and in vivo data have suggested that cross-reactivity exists only between compounds with very similar chemical structures. Sulfamethoxazole-specific lymphocytes may be activated by other antibacterial sulfonamides but not diuretics, antidiabetic drugs, or anti-COX2 NSAIDs with a sulfonamide group. Though it has been previously reported that 10% of patients with penicillin allergies will also develop allergic reactions to cephalosporin class antibiotics, the cross-reactivity is likely much lower, as is the incidence of true penicillin allergy itself, and severe reactions are very rare.

Recent data suggest that although the risk of developing a drug eruption to another drug is increased in persons with a prior reaction, “cross-sensitivity” is probably not the explanation. As an example, those with a history of an allergic-like reaction to penicillin are at greater risk of developing a reaction to antibacterial sulfonamides than to cephalosporins.

These data suggest that the list of drugs to avoid after a drug reaction should be limited to the causative one(s) and to a few very similar medications.

Because of growing evidence that some severe cutaneous reactions to drugs are associated with HLA genes, it is recommended that first-degree family

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ROLE OF TESTING FOR CAUSALITY AND DRUG RECHALLENGE

Recent data suggest that although the risk of developing a drug eruption to another drug is increased in persons with a prior reaction, "cross-reactivity" is not universal. These data suggest that the list of drugs to avoid after a drug reaction should be limited to the causative one(s) and to a few very similar medications.

Because of growing evidence that some severe cutaneous reactions to drugs are associated with HLA genes, it is recommended that first-degree family members of patients with severe cutaneous reactions also should avoid causative agents. This may be most relevant for sulfonamides and antiepileptic medications.

ROLE OF TESTING FOR CAUSALITY AND DRUG RECHALLENGE

The usefulness of laboratory tests, skin-prick, or patch testing to determine causality is debated and may be of limited practical value. Many in vitro immunologic assays have been developed for research purposes; however, the predictive value of these tests has not been validated in large series of affected patients. In some cases, diagnostic rechallenge may be appropriate, even for drugs with high rates of adverse reactions.

Skin-prick testing has clinical value in specific settings. In patients with a history suggesting immediate IgE-mediated reactions to penicillin, skin-prick testing with penicillins or cephalosporins has proven useful for identifying patients at risk of anaphylactic reactions to these agents. Negative skin tests do not totally rule out IgE-mediated reactivity; however, the risk of anaphylaxis in response to penicillin administration in patients with negative skin tests is about 1%. In contrast, two-thirds of patients with a positive skin test experience an allergic response upon rechallenge. The skin tests themselves carry a small risk of anaphylaxis.

For patients with delayed-type hypersensitivity, the clinical utility of skin tests remains questionable. At least one of a combination of several tests (prick, patch, and intradermal) is positive in 50–70% of patients with a reaction "definitely" attributed to a single medication. This low sensitivity corresponds to the observation that readministration of drugs with negative skin testing results in eruptions in 17% of cases.

Desensitization can be considered in those with a history of reaction to a medication that must be used again. Efficacy of such procedures has been demonstrated in cases of immediate reaction to penicillin and positive skin tests, anaphylactic reactions to platinum chemotherapy, and delayed reactions to sulfonamides in patients with AIDS. Desensitization is often successful in HIV-infected patients with morbilliform eruptions to sulfonamides but is not recommended in HIV-infected patients who developed erythroderma or a bullous reaction in response to prior sulfonamide exposure. Various protocols are available, including oral and parenteral approaches. Oral desensitization appears to have a lower risk of serious anaphylactic reaction. Desensitization carries the risk of anaphylaxis regardless of how it is performed and should be performed in monitored clinical settings such as an intensive care unit. After desensitization, many patients experience non-life-threatening reactions during therapy with the culprit drug.

REPORTING

Any severe reaction to drugs should be reported to a regulatory agency or to pharmaceutical companies. Because severe reactions are too rare to be detected in premarketing clinical trials, spontaneous reports are of critical importance for early detection of unexpected life-threatening events. To be useful, the report should contain enough details to permit ascertainment of severity and drug causality.

ACKNOWLEDGMENTS

We acknowledge the contribution of Drs. Jean-Claude Roujeau and Robert S. Stern to this chapter in previous editions.

FURTHER READING

Alfirevic A et al: Genetic testing for prevention of severe drug-induced skin rash. Cochrane Database Syst Rev 7:CD010891, 2019. [PubMed: 31314143]

Cornejo-Garcia JA et al: The genetics of drug hypersensitivity reactions. J Investig Allergol Clin Immunol 26:222, 2016. [PubMed: 27375032]

Duong TA et al: Severe cutaneous adverse reactions to drugs. Lancet 390:1996, 2017. [PubMed: 28476287]

Ko TM et al: Use of HLA-B*5801 genotyping to prevent allopurinol induced severe cutaneous adverse reactions in Taiwan: National prospective cohort study. BMJ 351:h4848, 2015. [PubMed: 26399967]

Lee S et al: Association of dipeptidyl peptidase 4 inhibitor use with risk of bullous pemphigoid in patients with diabetes. JAMA Dermatol 155:172, 2018. [PubMed: 29318313]

Oussalah A et al: Genetic variants associated with drug-induced immediate hypersensitivity reactions: A PRISMA-compliant systematic review. Allergy

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Lee S et al: Association of dipeptidyl peptidase 4 inhibitor use with risk of bullous pemphigoid in patients with diabetes. JAMA Dermatol 155:172, 2018.

Mayorga C et al: In vitro tests for drug hypersensitivity reactions: An ENDA/EAACI Drug Allergy Interest Group position paper. Allergy 71:1103, 2016. [PubMed: 26991315]

Oussalah A et al: Genetic variants associated with drug-induced immediate hypersensitivity reactions: A PRISMA-compliant systematic review. Allergy 71:443, 2016. [PubMed: 26678823]

Peter JG et al: Severe delayed cutaneous and systemic reactions to drugs: A global perspective on the science and art of current practice. J Allergy Clin Immunol Pract 5:547, 2017. [PubMed: 28483310]

Petrelli F et al: Antibiotic prophylaxis for skin toxicity induced by antiepidermal growth factor receptor agents: A systematic review and meta-analysis. Br J Dermatol 175:1166, 2016. [PubMed: 27214209]

Sassolas B et al: ALDEN, an algorithm for assessment of drug causality in Stevens-Johnson syndrome and toxic epidermal necrolysis: Comparison with case-control analysis. Clin Pharmacol Ther 88:60, 2010. [PubMed: 20375998]

Seminario-Vidal L et al: Society of Dermatology Hospitalists supportive care guidelines for the management of Stevens-Johnson syndrome/toxic epidermal necrolysis in adults. J Am Acad Dermatol 82:1553, 2020. [PubMed: 32151629]

Simonsen A et al: Cutaneous adverse reactions to anti-PD-1 treatment: A systematic review. J Am Acad Dermatol 83:1415, 2020. [PubMed: 32320766]

Zimmermann S et al: Systemic immunomodulating therapies for Stevens-Johnson syndrome and toxic epidermal necrolysis: A systematic review and meta-analysis. JAMA Dermatol 153:514, 2017. [PubMed: 28329382]

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 60: Cutaneous Drug Reactions, Robert G. Micheletti; Misha Rosenbach; Bruce U. Wintroub; Kanade Shinkai
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 61: Photosensitivity and Other Reactions to Sunlight

Alexander G. Marneros; David R. Bickers

SOLAR RADIATION

Sunlight is the most visible and obvious source of comfort in the environment. The sun provides the beneficial effects of warmth and vitamin D synthesis. However, acute and chronic sun exposure also has pathologic consequences. Cutaneous exposure to sunlight is a major cause of human skin cancer and can have immunosuppressive effects as well.

The sun's energy reaching the Earth's surface is limited to components of the ultraviolet (UV) spectrum, the visible spectrum, and portions of the infrared spectrum. The cutoff at the short end of the UV spectrum at ~290 nm is due primarily to stratospheric ozone—formed by highly energetic ionizing radiation—that prevents penetration to the earth's surface of the shorter, more energetic, potentially more harmful wavelengths of solar radiation. Indeed, concern about destruction of the ozone layer by chlorofluorocarbons released into the atmosphere has led to international agreements to reduce production of those chemicals.

Measurements of solar flux showed a 20-fold regional variation in the amount of energy at 300 nm that reaches the earth's surface. This variability relates to seasonal effects, the path that sunlight traverses through ozone and air, the altitude (a 4% increase for each 300 m of elevation), the latitude (increasing intensity with decreasing latitude), and the amount of cloud cover, fog, and pollution.

The major components of the photobiologic action spectrum that can affect human skin include the UV and visible wavelengths between 290 and 700 nm. In addition, the wavelengths beyond 700 nm in the infrared spectrum primarily emit heat and in certain circumstances may exacerbate the pathologic effects of energy in the UV and visible spectra.

The UV spectrum reaching the Earth represents <10% of total incident solar energy and is arbitrarily divided into two major segments, UV-B and UV-A, which constitute the wavelengths from 290 to 400 nm. UV-B consists of wavelengths between 290 and 320 nm. This portion of the photobiologic action spectrum is the most efficient in producing redness or erythema in human skin and thus is sometimes known as the "sunburn spectrum." UV-A includes wavelengths between 320 and 400 nm and is ~1000-fold less efficient in producing skin redness than is UV-B.

The wavelengths between 400 and 700 nm are visible to the human eye. The photon energy in the visible spectrum is not capable of damaging human skin in the absence of a photosensitizing chemical. Without the absorption of energy by a molecule, there can be no photosensitivity. Thus, the *absorption spectrum* of a molecule is defined as the range of wavelengths it absorbs, whereas the *action spectrum* for an effect of incident radiation is defined as the range of wavelengths that evoke the response.

Photosensitivity occurs when a photon-absorbing chemical (*chromophore*) present in the skin absorbs incident energy, becomes excited, and transfers the absorbed energy to various structures or to molecular oxygen.

UV RADIATION (UVR) AND SKIN STRUCTURE AND FUNCTION

Human skin consists of two major compartments: the outer epidermis, which is a stratified squamous epithelium, and the underlying dermis, which is rich in matrix proteins such as collagens and elastin. Both compartments are susceptible to damage from sun exposure. The epidermis and the dermis contain several chromophores capable of absorbing incident solar energy, including nucleic acids, proteins, and lipids. The outermost epidermal layer, the stratum corneum, is a major absorber of UV-B, and <10% of incident UV-B wavelengths penetrate through the epidermis to the dermis. Approximately 3% of radiation below 300 nm, 20% of radiation below 360 nm, and 33% of short visible radiation reach the basal cell layer in untanned human skin. UV-A readily penetrates to the dermis and is capable of altering structural and matrix proteins that contribute to photoaging of chronically sun-exposed skin, particularly in individuals of light complexion. Thus, longer wavelengths can penetrate more deeply into the skin.

Molecular Targets for UVR-Induced Skin Effects

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

structural changes are potentially mutagenic and are found in nonmelanoma skin cancers (NMSCs), including basal cell carcinoma (BCC), squamous

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Molecular Targets for UVR-Induced Skin Effects

Epidermal DNA—predominantly in keratinocytes and in Langerhans cells (dendritic antigen-presenting cells)—absorbs UV-B and undergoes structural changes between adjacent pyrimidine bases (thymine or cytosine), including the formation of cyclobutane dimers and 6,4-photoproducts. These structural changes are potentially mutagenic and are found in nonmelanoma skin cancers (NMSCs), including basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and Merkel cell carcinoma (MCC). They can be repaired by cellular mechanisms that result in their recognition and excision and the restoration of normal base sequences. The efficient repair of these structural aberrations is crucial, since individuals with defective DNA repair are at high risk for the development of cutaneous cancer. For example, patients with xeroderma pigmentosum, an autosomal recessive disorder, have a variably deficient repair of UV-induced photoproducts. The skin of these patients often shows the dry, leathery appearance of prematurely photoaged skin, and these patients have an increased frequency of skin cancer already in the first two decades of life. Studies in transgenic mice have verified the importance of functional genes that regulate these repair pathways in preventing the development of UV-induced skin cancer. DNA damage to Langerhans cells may also contribute to the known immunosuppressive effects of UV-B (see "Photoimmunology," later).

In addition to DNA, molecular oxygen is a target for incident solar UVR, leading to the generation of reactive oxygen species (ROS). These ROS can damage skin components through oxidative damage to DNA, oxidation of polyunsaturated fatty acids in lipids (lipid peroxidation), or oxidation of amino acids in proteins, or they can lead to oxidative deactivation of specific enzymes. UVR can also promote increased cross-linking and degradation of dermal matrix proteins and accumulation of abnormal dermal elastin, leading to photoaging changes known as *solar elastosis*.

Cutaneous Optics and Chromophores

Chromophores are endogenous or exogenous chemicals that can absorb physical energy. Endogenous chromophores are of two types: (1) normal components of skin, including nucleic acids, proteins, lipids, and 7-dehydrocholesterol (the precursor of vitamin D); and (2) components that are synthesized elsewhere in the body and that circulate in the bloodstream and diffuse into the skin, such as porphyrins. Normally, only trace amounts of porphyrins are present in the skin, but, in selected diseases known as the *porphyrias* **(Chap. 416)**, porphyrins are released into the circulation in increased amounts from the bone marrow and/or the liver and are transported to the skin, where they absorb incident energy both in the Soret band (~400 nm; short visible) and, to a lesser extent, in the red portion of the visible spectrum (580–660 nm). This energy absorption results in the generation of ROS that can mediate structural damage to the skin, manifested as erythema, edema, urticaria, or blister formation. It is of interest that photoexcited porphyrins are currently used in the treatment of BCCs and SCCs and their precursor lesions, actinic keratoses. Known as *photodynamic therapy* (PDT), this modality generates ROS in the skin, leading to cell death. Topical photosensitizers used in PDT are the porphyrin precursors 5-aminolevulinic acid and methyl aminolevulinate, which are readily converted to porphyrins in the skin. It is believed that PDT targets tumor cells for destruction more selectively than it targets adjacent nonneoplastic cells. The efficacy of such therapy requires appropriate timing of the application of methyl aminolevulinate or 5-aminolevulinic acid to the affected skin followed by exposure to artificial sources of visible light. High-intensity blue light has been used successfully for PDT of thin actinic keratoses. Red light PDT penetrates more deeply into the skin and is more beneficial in the treatment of superficial BCCs.

Acute Effects of Sun Exposure

The acute effects of skin exposure to sunlight include sunburn and vitamin D synthesis.

SUNBURN

This painful skin condition is an acute inflammatory response of the skin, predominantly to UV-B. Generally, an individual's ability to tolerate sunlight is inversely proportional to that individual's degree of melanin pigmentation. Melanin, a complex polymer of tyrosine derivatives, is synthesized in specialized epidermal dendritic cells known as *melanocytes* and is packaged into *melanosomes* that are transferred via dendritic processes into *keratinocytes*, thereby providing photoprotection (dissipating the vast majority of absorbed UVR in the skin) and simultaneously darkening the skin. Sun-induced melanogenesis is a consequence of increased tyrosinase activity in melanocytes. Central to the suntan response is the melanocortin-1 receptor (*MC1R*), and mutations in this gene contribute to the wide variation in human skin and hair color; individuals with red hair and fair skin typically have low MC1R activity. In the skin, there are two main types of melanin: eumelanin (providing brown and black pigmentation associated with high MC1R activity) and pheomelanin (providing red pigmentation associated with low MC1R activity). Pheomelanin is a cysteine-containing red polymer of benzothiazine units and has much weaker shielding capacity against UVR compared to eumelanin. This may explain why individuals with a higher proportion of pheomelanin (red hair/fair skin appearance) have an increased risk of melanoma formation. In addition, pheomelanin may also promote melanoma formation through induction of oxidative damage by amplifying UV-A–induced ROS but also through UVR-independent

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

mainly by keratinocytes in response to UVR. The UV-induced expression of this hormone is controlled by the tumor suppressor p53, and absence of functional p53 attenuates the tanning response. Activation of the melanocortin receptor leads to increased intracellular cyclic adenosine 5′-

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

higher proportion of pheomelanin (red hair/fair skin appearance) have an increased risk of melanoma formation. In addition, pheomelanin may also promote melanoma formation through induction of oxidative damage by amplifying UV-A–induced ROS but also through UVR-independent mechanisms.

The human *MC1R* gene encodes a G protein–coupled receptor that binds α-melanocyte-stimulating hormone (α-MSH), which is secreted in the skin mainly by keratinocytes in response to UVR. The UV-induced expression of this hormone is controlled by the tumor suppressor p53, and absence of functional p53 attenuates the tanning response. Activation of the melanocortin receptor leads to increased intracellular cyclic adenosine 5′-monophosphate (cAMP) and protein kinase A activation, resulting in an increased transcription of the microphthalmia-associated transcription factor (MITF), which stimulates melanogenesis. Since the precursor of α-MSH, proopiomelanocortin produced by keratinocytes, is also the precursor of β-endorphins, UVR may result in not only increased pigmentation but also increased β-endorphin production in the skin, an effect that has been hypothesized to promote sun-seeking behaviors and even mediate addiction to tanning.

The Fitzpatrick classification of human skin phototypes is based on the efficiency of the epidermal-melanin unit, which usually can be ascertained by asking an individual two questions: (1) Do you burn after sun exposure? (2) Do you tan after sun exposure? The answers to these questions permit division of the population into six skin types, varying from type I (always burn, never tan) to type VI (never burn, always tan) **(Table 61-1)**.

TABLE 61-1

Skin Type and Sunburn Sensitivity (Fitzpatrick Classification)

TYPE	DESCRIPTION
I	Always burn, never tan
II	Always burn, sometimes tan
III	Sometimes burn, sometimes tan
IV	Sometimes burn, always tan
V	Never burn, sometimes tan
VI	Never burn, always tan

Sunburn erythema is due to vasodilation of dermal blood vessels. There is a lag time (usually 4–12 h) between skin exposure to sunlight and the development of visible redness. The action spectrum for sunburn erythema includes UV-B and UV-A, although UV-B is much more efficient than UV-A in evoking the response. However, UV-A may contribute to sunburn erythema at midday, when much more UV-A than UV-B is present in the solar spectrum. The erythema that accompanies the inflammatory response induced by UVR results from the orchestrated release of cytokines along with growth factors and the generation of ROS. Furthermore, UV-induced activation of nuclear factor κB–dependent gene transcription can augment release of several proinflammatory cytokines and vasoactive mediators. These cytokines and mediators accumulate locally in sunburned skin, providing chemotactic factors that attract neutrophils, macrophages, and T lymphocytes, which promote the inflammatory response. UVR also stimulates infiltration of inflammatory cells through induced expression of adhesion molecules such as E-selectin and intercellular adhesion molecule 1 on endothelial cells and keratinocytes. UVR has been shown to activate phospholipase A_2, resulting in increases in eicosanoids such as prostaglandin E_2, which is known to be a potent inducer of sunburn erythema. The role of eicosanoids in this reaction has been verified by studies showing that nonsteroidal anti-inflammatory drugs (NSAIDs) can reduce sunburn erythema.

Epidermal changes in sunburn include the induction of “sunburn cells,” which are keratinocytes undergoing p53-dependent apoptosis as a defense, with elimination of cells that harbor UV-B–induced structural DNA damage.

VITAMIN D SYNTHESIS AND PHOTOCHEMISTRY

Cutaneous exposure to UV-B causes photolysis of epidermal 7-dehydrocholesterol, converting it to pre–vitamin D_3, which then undergoes

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

exploited therapeutically in psoriasis with the topical application of synthetic vitamin D analogues. In addition, vitamin D is increasingly thought to have beneficial effects in several other inflammatory conditions, and some evidence suggests that—besides its classic physiologic effects on calcium

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Cutaneous exposure to UV-B causes photolysis of epidermal 7-dehydrocholesterol, converting it to pre–vitamin D_3, which then undergoes temperature-dependent isomerization to form the stable hormone vitamin D_3. This compound diffuses to the dermal vasculature and circulates to the liver and kidney, where it is converted to the dihydroxylated functional hormone 1,25-dihydroxyvitamin D_3. Vitamin D metabolites from the circulation and those produced in the skin itself can augment epidermal differentiation signaling and inhibit keratinocyte proliferation. These effects are exploited therapeutically in psoriasis with the topical application of synthetic vitamin D analogues. In addition, vitamin D is increasingly thought to have beneficial effects in several other inflammatory conditions, and some evidence suggests that—besides its classic physiologic effects on calcium metabolism and bone homeostasis—it is associated with a reduced risk of various internal malignancies. There is controversy regarding the risk-to-benefit ratio of sun exposure for vitamin D homeostasis. At present, it is important to emphasize that no clear-cut evidence suggests that the use of sunscreens substantially diminishes vitamin D levels. Since aging also substantially decreases the ability of human skin to photocatalytically produce vitamin D_3, the widespread use of sunscreens that filter out UV-B has led to concerns that the elderly might be unduly susceptible to vitamin D deficiency. However, the amount of sunlight needed to produce sufficient vitamin D is small and does not justify the risks of skin cancer and other types of photodamage linked to increased sun exposure or tanning behavior. Nutritional supplementation of vitamin D is a preferable strategy for patients with vitamin D deficiency.

Chronic Effects of Sun Exposure: Nonmalignant

The clinical features of photoaging (*dermatoheliosis*) consist of wrinkling, blotchiness, and telangiectasia, as well as a roughened, irregular, "weather-beaten" leathery appearance.

UVR is important in the pathogenesis of photoaging in human skin, and ROS are likely involved. The dermis and its connective tissue matrix are major targets for sun-associated chronic damage that manifests as solar elastosis, a massive increase in thickened irregular masses of abnormal-appearing elastic fibers. Collagen fibers are also abnormally clumped in the deeper dermis of sun-damaged skin. The chromophores, the action spectra, and the specific biochemical events orchestrating these changes are only partially understood, although more deeply penetrating UV-A seems to be primarily involved. Chronologically aged sun-protected skin and photoaged skin share important molecular features, including connective tissue damage and elevated levels of matrix metalloproteinases (MMPs). MMPs are enzymes involved in the degradation of the extracellular matrix. UV-A induces expression of some MMPs, including MMP-1 and MMP-3, leading to increased collagen breakdown. In addition, UV-A reduces type I procollagen messenger RNA (mRNA) expression. Thus, chronic UVR alters the structure and function of dermal collagen both by inhibiting its synthesis and enhancing its breakdown. Based on these observations, it is not surprising that high-dose UV-A phototherapy may have beneficial effects in some patients with localized fibrotic diseases of the skin, such as localized scleroderma.

Chronic Effects of Sun Exposure: Malignant

One of the major known consequences of chronic excessive skin exposure to sunlight is NMSC, including SCCs, BCCs and MCCs **(Chap. 76)**. A model for skin cancer induction involves three major steps: initiation, promotion, and progression. Exposure of human skin to sunlight results in *initiation*, a step by which structural (mutagenic) changes in DNA evoke an irreversible alteration in the target cell (keratinocyte) that begins the tumorigenic process. Exposure to a tumor initiator such as UV-B is believed to be a necessary but not a sufficient step in the malignant process, since initiated skin cells not exposed to tumor promoters generally do not develop into tumors. The second stage in tumor development is *promotion*, a multistep process by which chronic exposure to sunlight evokes further changes that culminate in the clonal expansion of initiated cells and cause the development of premalignant growths known as *actinic keratoses*, which may progress to form SCCs. As a result of extensive studies, it seems clear that UV-B is a *complete carcinogen*, meaning that it can act as both a tumor initiator and a tumor promoter. The third and final step in the malignant process is *malignant conversion* of benign precursors into cancers, a process thought to enhance genetic instability.

On a molecular level, skin carcinogenesis results from the accumulation of gene mutations that cause inactivation of tumor suppressors, activation of oncogenes, or reactivation of cellular signaling pathways that normally are expressed only during embryologic epidermal development that drive cell proliferation. Interestingly, a large number of UV-induced oncogenic driver mutations that are present in SCCs can already be found in aged sun-exposed normal skin, leading to a growth advantage and innumerable precancerous clones carrying cancer-causing mutations. These mutations occur particularly often in genes that affect proliferation of epidermal stem cells (e.g., NOTCH receptor genes). The pattern of oncogenic gene mutations in aged sun-exposed skin shows considerable overlap with the mutations identified in SCCs, while there is little overlap with the mutations identified in BCCs or melanomas. For example, ~20% of normal aged sun-exposed skin cells and ~60% of SCCs carry driver mutations in *NOTCH1*. Additionally, the accumulation of mutations in the tumor-suppressor gene *p53* can also promote skin carcinogenesis. Indeed, the majority of both human and murine UV-induced skin cancers have characteristic UVR-induced *p53* mutations (C → T and CC → TT transitions). Studies in mice have shown that sunscreens substantially reduce the frequency of these signature mutations in *p53* and inhibit the induction of tumors. These comparisons of UVR-induced oncogenic driver mutations per cell than cells in aged sun-exposed normal skin. Furthermore, while aged sun-exposed skin and SCCs carry similar UVR-induced mutations in *p53* or NOTCH receptors, oncogenic mutations in other genes (e.g., *CDKN2A*) were mainly found in SCCs and not in aged

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

particularly often in genes that affect proliferation of epidermal stem cells (e.g., NOTCH receptor genes). The pattern of oncogenic gene mutations in UV-induced skin cancers have characteristic UVR-induced *p53* mutations (C → T and CC → TT transitions). Studies in mice have shown that sunscreens can substantially reduce the frequency of these signature mutations in *p53* and inhibit the induction of tumors. The comparison of UVR-induced gene mutations between aged sun-exposed normal skin and SCCs supports the hypothesis of a progressive accumulation of additional oncogenic mutations that eventually lead to the transition from precancerous cell clones to SCCs. It has been estimated that SCCs harbor ~10 times more oncogenic driver mutations per cell than cells in aged sun-exposed normal skin. Furthermore, while aged sun-exposed skin and SCCs carry similar UVR-induced mutations in *p53* or NOTCH receptors, oncogenic mutations in other genes (e.g., *CDKN2A*) were mainly found in SCCs and not in aged sun-exposed skin, which are thus likely to play a critical role in malignant progression.

Compared to SCCs, BCCs carry a distinct mutational profile in specific genes. BCCs harbor inactivating mutations particularly in the tumor-suppressor gene *patched* or activating mutations in the oncogene *smoothened*, which result in the constitutive activation of the sonic hedgehog signaling pathway and increased cell proliferation. There is also evidence linking alterations in the Wnt/β-catenin signaling pathway, which is known to be critical for hair follicle development, to skin cancer as well. Thus, interactions between this pathway and the hedgehog signaling pathway appear to be involved in both skin carcinogenesis and embryologic development of the skin and hair follicles.

Clonal analysis in mouse models of BCC revealed that tumor cells arise from stem cells of the interfollicular epidermis and the upper infundibulum of the hair follicle. These BCC-initiating cells are reprogrammed to resemble embryonic hair follicle progenitors, whose tumor-initiating ability depends on activation of the Wnt/β-catenin signaling pathway.

SCC initiation occurs both in the interfollicular epidermis and in the hair follicle bulge stem cell populations. In mouse models, the combination of mutant K-Ras and p53 is sufficient to induce invasive SCCs from these cell populations.

The transcription factor Myc is important for stem cell maintenance in the skin, and oncogenic activation of Myc has been implicated in the development of BCCs and SCCs.

The third NMSC is MCC, which is named after its resemblance to Merkel cells in the skin. The incidence of MCCs has been increasing in recent years for unknown reasons. The age-adjusted global incidence is about 1 in 100,000. Just like SCC and BCC, patients with MCCs are usually fair-skinned males in the sixth to eighth decades of life who are living in geographic regions with greater solar UVR. These tumors occur predominantly on the head and neck in older individuals and on the trunk in younger people. MCCs also have a higher incidence among immunosuppressed patients. MCCs are aggressive and life-threatening, poorly differentiated neuroendocrine carcinomas. Overall survival at 5 years is around 50% for local disease, 35% for nodal disease, and 15% for metastatic disease. While the majority of MCCs present locally, nodal and metastatic disease can occur simultaneously. The pathogenesis of MCCs is closely connected to the Merkel cell polyoma virus (MCPyV). It is now recognized that MCCs can either be MCPyV positive or MCPyV negative. MCPyV-negative MCCs manifest high levels of classic UV-induced signature mutations (C to T or CC to TT) and inactivation of tumor suppressor genes, which could explain the growth of these viral-negative lesions. MCPyV-positive tumors are thought to grow secondary to viral integration into the host genome and acquisition of a truncating mutation of the large T antigen that results in the production of viral oncoproteins. The growth of MCPyV-positive tumors may be further promoted by UVR-induced local immunosuppression. Both forms of MCC are immunogenic, and metastatic MCC has been treated in some patients successfully with PD-1/PD-L1 immune checkpoint inhibitors.

In summary, NMSC involves mutations and alterations in multiple genes and pathways that occur as a result of their chronic accumulation driven by exposure to environmental factors such as solar UVR.

Epidemiologic studies have linked excessive sun exposure to an increased risk of NMSCs and melanoma of the skin; the evidence is far more direct for NMSCs (BCCs, SCCs, and MCCs) than for melanoma. Approximately 80% of NMSCs develop on sun-exposed body areas, including the face, neck, and hands. Major risk factors include male sex, childhood sun exposures, older age, fair skin, and residence at latitudes relatively close to the equator. Individuals with darker-pigmented skin have a lower risk of skin cancer than do fair-skinned individuals. More than 2 million individuals in the United States develop NMSC annually, and the lifetime risk that a fair-skinned individual will develop such a neoplasm is estimated at ~15%. The incidence of NMSC in the population is increasing at a rate of 2–3% per year, likely due to earlier detection and increased opportunities for outdoor activities.

The relationship of sun exposure to melanoma development is less direct, but strong evidence supports an association. Clear-cut risk factors include a positive family or personal history of melanoma and multiple dysplastic nevi. Melanomas can occur during adolescence; the implication is that the latent period for tumor growth is shorter than that for NMSC. For reasons that are only partially understood, melanomas are among the most rapidly increasing human malignancies (Chap. 76). One potential explanation is the widespread use of indoor tanning. It is estimated that 30 million people tan indoors in the United States annually, including >2 million adolescents. Furthermore, epidemiologic studies suggest that life in a sunny climate from birth or early childhood may increase the risk of melanoma development. In general, risk does not correlate with cumulative sun exposure but may be related to the duration and extent of exposure in childhood

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

in mouse models showing that pheomelanin is less efficient in protecting against melanoma than is eumelanin and may promote melanoma through UVR-independent mechanisms

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

from birth or early childhood may increase the risk of melanoma development. In general, risk does not correlate with cumulative sun exposure but may be related to the duration and extent of exposure in childhood.

However, in contrast to NMSCs, melanoma frequently develops in non–sun-exposed skin, and oncogenic mutations in melanoma may also not be UVR-signature mutations. These observations suggest that UVR-independent factors may contribute to melanomagenesis, which is consistent with findings in mouse models showing that pheomelanin is less efficient in protecting against melanoma than is eumelanin and may promote melanoma through UVR-independent mechanisms.

Importantly, mutations in BRAF and NRAS that lead to activation of a growth-promoting signaling cascade are frequently found in melanoma (but not in SCCs or BCCs), which has led to the development of specific inhibitors of this pathway for the treatment of BRAF-mutant melanoma. However, a high mutational load in melanoma may not be equated with a more unfavorable prognosis. Tumor-specific missense mutations in melanomas can result in neoantigens that facilitate an immune response to the tumor cell. A major advance in treating melanoma, termed immune checkpoint blockade, targets inhibitors of cytotoxic T effector function. For example, the PD-1/PD-L1 interaction inhibits tumor cell apoptosis, promotes peripheral T effector cell exhaustion, and induces conversion of T effector cells to regulatory T cells. Checkpoint inhibitor treatment (e.g., with antibodies that inhibit PD-1 or PD-L1) disrupts this interaction and has resulted in a durable and potent immune destruction of melanoma cells in a subset of patients, leading to prolonged survival of patients with locally advanced or metastatic melanoma. It has recently been shown that a high mutational load in melanomas correlates with improved therapeutic outcome to immune checkpoint blockade, consistent with the hypothesis that acquired missense mutations in the tumor cells lead to neoantigens that increase the vulnerability of these melanoma cells to attack by activated T cells.

GLOBAL CONSIDERATIONS

The frequency of skin cancer shows strong geographic variation, depending on the skin phototype of the majority of the population in these geographic areas, but also depending on the intensity of UVR. For example, both melanoma and NMSCs are particularly common in Australia.

Photoimmunology

Exposure to solar radiation causes both local and systemic immunosuppression and involves both the innate and adaptive immune systems. Local immunosuppression is defined as inhibition of immune responses to antigens applied at the irradiated site, whereas systemic immunosuppression is defined as inhibition of immune responses to antigens applied at remote, unirradiated sites. An example of local immunosuppression is that human skin exposure to modest doses of UV-B can deplete the epidermal antigen-presenting Langerhans cells, thereby reducing the degree of allergic sensitization to topical application of the potent contact allergen dinitrochlorobenzene at the irradiated skin site. An example of the systemic immunosuppressive effects of higher doses of UVR is the diminished immunologic response to antigens introduced either epicutaneously or intracutaneously at sites remote from the irradiated site.

The major chromophores in the upper epidermis that are known to initiate UV-mediated immunosuppression include DNA, *trans*-urocanic acid, and membrane components. The action spectrum for UV-induced immunosuppression closely mimics the absorption spectrum of DNA. UVR-induced cyclobutane pyrimidine dimers in Langerhans cells may inhibit antigen presentation. The absorption spectrum of epidermal urocanic acid closely mimics the action spectrum for UV-B–induced immunosuppression. Urocanic acid is a metabolic product of the essential amino acid histidine and accumulates in the upper epidermis through breakdown of the histidine-rich protein filaggrin due to the absence of its catabolizing enzyme in keratinocytes. Urocanic acid is synthesized as a *trans*-isomer, and UV-induced *trans-cis* isomerization of urocanic acid in the stratum corneum drives immunosuppression. *Cis*-urocanic acid may exert its immunosuppressive effects through a variety of mechanisms, including inhibition of antigen presentation by Langerhans cells.

Various additional immunomodulatory factors and cytokines have been implicated in UVR-induced systemic immunosuppression, including tumor necrosis factor-α, interleukin 4 (IL-4), interleukin 10 (IL-10), and eicosanoids. Keratinocytes can release multiple immunomodulators as a response to UVR-induced cell damage that result in an immunosuppressive environment. Induction of IL-4–producing natural killer T cells and of regulatory T cells and B cells has been linked to cell-mediated and humoral immunosuppression as a consequence of UVR damage to skin. Moreover, UVR-induced formation of damage-associated molecular patterns (DAMPs) from necrotic keratinocytes can lead to a type I interferon innate immune response via activation of Toll-like receptor signaling.

One important consequence of chronic sun exposure and associated immunosuppression is an enhanced risk of skin cancer. In part, UV-B activates regulatory T cells that suppress antitumor immune responses via IL-10 expression, whereas in the absence of high UV-B exposure, epidermal Langerhans cells present tumor-associated antigens and induce protective immunity, thereby inhibiting skin tumorigenesis. UV-induced DNA damage

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

BCCs and SCCs, and these skin cancers are the most common types of malignancies arising in these patients. The important contributory role of UVR for the formation of these skin cancers in immunosuppressed individuals is highlighted by the observation that nonwhite transplant recipients

Langerhans cells present tumor-associated antigens and induce protective immunity, thereby inhibiting skin tumorigenesis. UV-induced DNA damage is a major molecular trigger of this immunosuppressive effect.

Perhaps the most graphic demonstration of the role of long-term immunosuppression in enhancing the risk of NMSC comes from studies of organ transplant recipients who require lifelong immunosuppressive/antirejection drug regimens. More than 50% of organ transplant recipients develop BCCs and SCCs, and these skin cancers are the most common types of malignancies arising in these patients. The important contributory role of UVR for the formation of these skin cancers in immunosuppressed individuals is highlighted by the observation that nonwhite transplant recipients develop these skin cancers far less often than white transplant recipients. Rates of BCC and SCC increase with the duration and degree of immunosuppression. Transplant recipients ideally should be screened prior to organ transplantation, be monitored closely thereafter, and adhere to rigorous photoprotection measures, including the use of sunscreens and protective clothing as well as sun avoidance. Notably, immunosuppressive drugs that target the mTOR pathway, such as sirolimus and everolimus, may reduce the risk of NMSC in organ transplant recipients compared to that associated with the use of calcineurin inhibitors (cyclosporine and tacrolimus). The latter may contribute to NMSC formation not only through their immunosuppressive effects but also through suppression of p53-dependent cancer cell senescence pathways independent of host immunity.

Whereas the immunosuppressive effects of UVR contribute to skin cancer, UVR can also exacerbate autoimmune and inflammatory diseases of the skin, including systemic lupus erythematosus (SLE). It has been proposed that in SLE UVR-induced damage to DNA may promote autoantibody formation.

PHOTOSENSITIVITY DISEASES

The diagnosis of photosensitivity requires elicitation of a careful history to define the duration of signs and symptoms, the length of time between exposure to sunlight and the development of subjective symptoms and visible changes in the skin. The age of onset can also be a helpful diagnostic clue. For example, the acute photosensitivity of erythropoietic protoporphyria (EPP) almost always begins in infancy or early childhood, whereas the chronic photosensitivity of porphyria cutanea tarda (PCT) typically begins in the fourth and fifth decades of life. A patient's history of exposure to topical and systemic drugs and chemicals may provide important diagnostic clues. Many classes of drugs can cause photosensitivity on the basis of either phototoxicity or photoallergy.

Examination of the skin may offer important clues. Anatomic areas that are naturally protected from direct sunlight, such as the hairy scalp, the upper eyelids, the retroauricular areas, and the infranasal and submental regions, may be spared, whereas exposed areas show characteristic features of the pathologic process. These anatomic localization patterns are often helpful, but not infallible, in making the diagnosis. For example, airborne contact sensitizers that are blown onto the skin may produce dermatitis that can be difficult to distinguish from photosensitivity despite the fact that such material may trigger skin reactivity in areas shielded from direct sunlight.

Many dermatologic conditions may be caused or aggravated by sunlight **(Table 61-2)**. The role of light in evoking these responses may be dependent on genetic abnormalities ranging from well-described defects in DNA repair that occur in xeroderma pigmentosum to the inherited abnormalities in heme synthesis that characterize the porphyrias.

TABLE 61-2

Classification of Photosensitivity Diseases

TYPE	DISEASE
Genetic	Erythropoietic porphyria
	Erythropoietic protoporphyria
	Porphyria cutanea tarda—familial
	Variegate porphyria
	Hepatoerythropoietic porphyria
	Albinism
	Rothmund-Thomson syndrome

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

	Albinism
	Xeroderma pigmentosum
	Rothmund-Thomson syndrome
	Bloom syndrome
	Cockayne syndrome
	Kindler syndrome
	Phenylketonuria
Metabolic	Porphyria cutanea tarda—sporadic
	Hartnup disease
	Kwashiorkor
	Pellagra
	Carcinoid syndrome
Phototoxic	
Internal	Drugs
External	Drugs, plants, food
Photoallergic	
Immediate	Solar urticaria
Delayed	Drug photoallergy
	Persistent light reaction/chronic actinic dermatitis
Neoplastic and degenerative	Photoaging
	Actinic keratosis
	Melanoma and nonmelanoma skin cancer
Idiopathic	Polymorphous light eruption
	Hydroa aestivale
	Actinic prurigo
Photoaggravated	Lupus erythematosus
	Systemic
	Discoid

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

	Systemic
	Subacute cutaneous
	Discoid
	Dermatomyositis
	Herpes simplex
	Lichen planus actinicus
	Acne vulgaris (aestivale)

Polymorphous Light Eruption

The most common type of photosensitivity disease is *polymorphous light eruption* (PMLE). Many affected individuals may never seek medical attention because the condition is often transient, becoming manifest in the spring with initial sun exposure but then subsiding spontaneously with continuing exposure, a phenomenon known as "hardening." The major manifestations of PMLE include (often intensely) pruritic erythematous papules that may coalesce into plaques in a patchy distribution on exposed areas of the trunk and forearms. The face is usually less affected. Whereas the morphologic skin findings remain similar for each patient with subsequent recurrences, significant interindividual variations in skin findings are characteristic (hence the term *polymorphous*).

A skin biopsy and phototest procedures in which skin is exposed to multiple erythemal doses of UV-A and UV-B may aid in the diagnosis. The action spectrum for PMLE is usually within these portions of the solar spectrum.

Whereas the treatment of an acute flare of PMLE may require topical or systemic glucocorticoids, approaches to preventing PMLE are important and include the use of high-SPF broad-spectrum sunscreens as well as the induction of "hardening" by the cautious administration of artificial UV-B (broad-band or narrow-band) and/or UV-A radiation or the use of psoralen plus UV-A (PUVA) photochemotherapy for ~4 weeks before initial sun exposure. Such prophylactic phototherapy or photochemotherapy at the beginning of spring may prevent the occurrence of PMLE throughout the summer.

Actinic prurigo is a photo-induced pruritic eruption that shares similarities with PMLE and often occurs in the spring; however, it can persist throughout the summer and extend into the winter months.

Phototoxicity and Photoallergy

These photosensitivity disorders are related to the topical or systemic administration of drugs and other chemicals that can act as chromophores. Both reactions require the absorption of energy by a drug or chemical with consequent production of an excited-state photosensitizer that can transfer its absorbed energy to a bystander molecule or to molecular oxygen, thereby generating tissue-destructive chemical species, including ROS.

Phototoxicity is a nonimmunologic reaction that can be caused by a broad range of drugs and chemicals, some of which are listed in **Table 61-3**. The usual clinical manifestations include erythema resembling a sunburn reaction that quickly desquamates, or "peels," within several days. In addition, edema, vesicles, and bullae may occur. A common phototoxic reaction that occurs after contact with plant-derived furocoumarins and exposure to UV-A radiation is called phytophotodermatitis.

TABLE 61-3

Drugs That May Cause a Phototoxic Reaction

DRUG	TOPICAL	SYSTEMIC
Fluoroquinolones		+

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 61-3

Drugs That May Cause a Phototoxic Reaction

DRUG	TOPICAL	SYSTEMIC
Amiodarone		+
Dacarbazine		+
Fluoroquinolones		+
5-Fluorouracil	+	+
Furosemide		+
Nalidixic acid		+
Phenothiazines		+
Psoralens	+	+
Retinoids	+/−	+
Sulfonamides		+
Sulfonylureas		+
Tetracyclines		+
Thiazides		+
Vinblastine		+

Photoallergy is much less common and is distinct in that it is an immunopathologic process. The excited-state photosensitizer may create highly unstable haptenic free radicals that bind covalently to macromolecules to form a functional antigen (photoallergen) capable of evoking a delayed-type hypersensitivity response. Most photoallergic reactions are initiated by UV-A rather than UV-B exposure. Some drugs and chemicals that can produce photoallergy are listed in **Table 61-4**. The clinical manifestations typically differ from those of phototoxicity in that an intensely pruritic eczematous dermatitis tends to predominate and evolves into lichenified, thickened, "leathery" changes in sun-exposed areas. A small subset (perhaps 5–10%) of patients with photoallergy may develop a persistent exquisite hypersensitivity to light even when the offending drug or chemical is identified and eliminated, a condition known as *persistent light reaction*.

TABLE 61-4

Drugs That May Cause a Photoallergic Reaction

DRUG	TOPICAL	SYSTEMIC
6-Methylcoumarin	+	
Aminobenzoic acid and esters	+	
Diclofenac		+

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

TABLE 61-4

Drugs That May Cause a Photoallergic Reaction

DRUG	TOPICAL	SYSTEMIC
6-Methylcoumarin	+	
Aminobenzoic acid and esters	+	
Bithionol	+	
Chlorpromazine		+
Diclofenac		+
Fluoroquinolones		+
Halogenated salicylanilides	+	
Hypericin (St. John's wort)	+	+
Musk ambrette	+	
Piroxicam		+
Promethazine		+
Sulfonamides		+
Sulfonylureas		+

An uncommon type of persistent photosensitivity is known as *chronic actinic dermatitis.* The affected patients are typically elderly men with a long history of preexisting allergic contact dermatitis or photosensitivity. Common photoallergens associated with this condition are sunscreen ingredients and plant photoallergens. These individuals are usually exquisitely sensitive to UV-B, UV-A, and visible wavelengths.

Phototoxicity and photoallergy often can be diagnostically confirmed by phototest procedures. In patients with suspected phototoxicity, determining the minimal erythemal dose (MED) while the patient is exposed to a suspected agent and then repeating the MED after discontinuation of the agent may provide a clue to the causative drug or chemical. Photopatch testing can be performed to confirm the diagnosis of photoallergy. In this simple variant of ordinary patch testing, a series of known photoallergens is applied to the skin in duplicate, and one set is irradiated with a suberythemal dose of UV-A. The development of eczematous changes at sites exposed to sensitizer and light is a positive result. The characteristic abnormality in patients with persistent light reaction is a diminished threshold to erythema evoked by UV-B. Patients with chronic actinic dermatitis usually manifest a broad spectrum of UV hyperresponsiveness and require meticulous photoprotection, including avoidance of sun exposure, use of high-SPF (>30) sunscreens, and, in severe cases, systemic immunosuppression, such as with azathioprine.

The management of drug photosensitivity involves first and foremost the elimination of exposure to the chemical agents responsible for the reaction and the minimization of sun exposure. The acute symptoms of phototoxicity may be ameliorated by cool moist compresses, topical glucocorticoids, and systemically administered NSAIDs. In severely affected individuals, a tapered course of systemic glucocorticoids may be useful. Judicious use of analgesics may be necessary.

Photoallergic reactions require a similar management approach. Furthermore, patients with persistent light reaction and chronic actinic dermatitis must be meticulously protected against light exposure. In selected patients to whom chronic systemic high-dose glucocorticoids pose unacceptable

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The porphyrias **(Chap. 416)** are a group of diseases that have in common inherited or acquired derangements in the synthesis of heme. Heme is an

Photoallergic reactions require a similar management approach. Furthermore, patients with persistent light reaction and chronic actinic dermatitis must be meticulously protected against light exposure. In selected patients to whom chronic systemic high-dose glucocorticoids pose unacceptable risks, it may be necessary to employ an immunosuppressive drug such as azathioprine, cyclophosphamide, cyclosporine, or mycophenolate mofetil.

Porphyria

The porphyrias **(Chap. 416)** are a group of diseases that have in common inherited or acquired derangements in the synthesis of heme. Heme is an iron-chelated tetrapyrrole or porphyrin, and only the nonmetal chelated porphyrins are potent photosensitizers that absorb light intensely in both the short (400–410 nm) and the long (580–650 nm) portions of the visible spectrum.

Heme cannot be reutilized and must be synthesized continuously. The two body compartments with the largest capacity for its production are the bone marrow and the liver. Accordingly, the porphyrias originate in one or the other of these organs, with an end result of excessive endogenous production of potent photosensitizing porphyrins. The porphyrins circulate in the bloodstream and diffuse into the skin, where they absorb solar energy, become photoexcited, generate ROS, and evoke cutaneous photosensitivity. The mechanism of porphyrin photosensitization is known to be photodynamic, or oxygen-dependent, and is mediated by ROS such as singlet oxygen and superoxide anions.

The group of cutaneous porphyrias can be classified as causing either (1) chronic blistering photosensitivity or (2) acute nonblistering photosensitivity. Chronic cutaneous porphyrias include PCT, congenital erythropoietic porphyria (CEP), hepatoerythropoietic porphyria (HEP), hereditary coproporphyria (HCP), and variegate porphyria (VP). CEP, HEP, and PCT manifest only with cutaneous symptoms, while HCP and VP have acute neurovisceral symptoms in addition to the skin photosensitivity. Acute cutaneous nonblistering porphyrias include EPP and X-linked protoporphyria (XLP). Representative examples of chronic and acute cutaneous porphyrias are discussed below.

Porphyria cutanea tarda (PCT) is the most common type of porphyria and is associated with decreased activity of the heme pathway enzyme uroporphyrinogen decarboxylase (UROD) to <20% of normal. Increased iron and various acquired factors (e.g., alcohol consumption, estrogens, smoking, hepatitis C or HIV infection) can reduce UROD activity. There are two basic types of PCT: (1) the sporadic or acquired type, generally seen in individuals ingesting ethanol or receiving estrogens; and (2) the inherited type, in which there is autosomal dominant transmission of deficient enzyme activity (resulting in heterozygosity for UROD with a reduction to 50% of UROD enzymatic activity and, thus, predisposing the individual to PCT). Both forms are associated with increased hepatic iron stores.

In both types of PCT, the predominant feature is chronic photosensitivity characterized by increased fragility of sun-exposed skin, particularly areas subject to repeated trauma such as the dorsa of the hands, the forearms, the face, and the ears. The predominant skin lesions are vesicles and bullae that rupture, producing moist erosions (often with a hemorrhagic base) that heal slowly, with crusting and purplish discoloration of the affected skin. Hypertrichosis, mottled pigmentary change, and scleroderma-like induration are associated features. The diagnosis can be confirmed biochemically by measurement of urinary porphyrin excretion, plasma porphyrin assay, and assay of erythrocyte and/or hepatic UROD. Multiple mutations of the *UROD* gene have been identified in human populations. Some patients with PCT have associated mutations in the *HFE* gene, which is linked to hemochromatosis and leads to increased iron absorption by reducing hepcidin expression; these mutations could contribute to the iron overload precipitating PCT, although iron status as measured by serum ferritin, iron levels, and transferrin saturation is no different from that in PCT patients without *HFE* mutations.

Treatment of PCT consists of repeated phlebotomies to diminish the excessive hepatic iron stores and/or intermittent (twice weekly) low doses of orally administered hydroxychloroquine. This treatment is highly effective for PCT but not suited for treatment of other porphyrias. Long-term remission of the disease can often be achieved if the patient eliminates exposure to porphyrinogenic agents, such as ethanol or estrogens, and avoids sun exposure.

Erythropoietic protoporphyria (EPP) is an acute nonblistering cutaneous porphyria, originates in the bone marrow, and is due to genetic mutations that in most cases decrease the activity of the mitochondrial enzyme ferrochelatase. The major clinical features include acute photosensitivity characterized by painful burning and stinging of exposed skin that often develops during or just after sun exposure. There may be associated skin swelling and, after repeated episodes, a waxlike scarring.

Detection of increased plasma protoporphyrin (PROTO) helps distinguish EPP from lead poisoning and iron-deficiency anemia, in both of which erythrocyte PROTO levels are elevated in the absence of cutaneous photosensitivity. This can be explained by the fact that metal-chelated PROTO is not a photosensitizer.

Rigorous sunlight protection is essential in the management of EPP. Notably, the U.S. Food and Drug Administration (FDA) has approved a synthetic

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

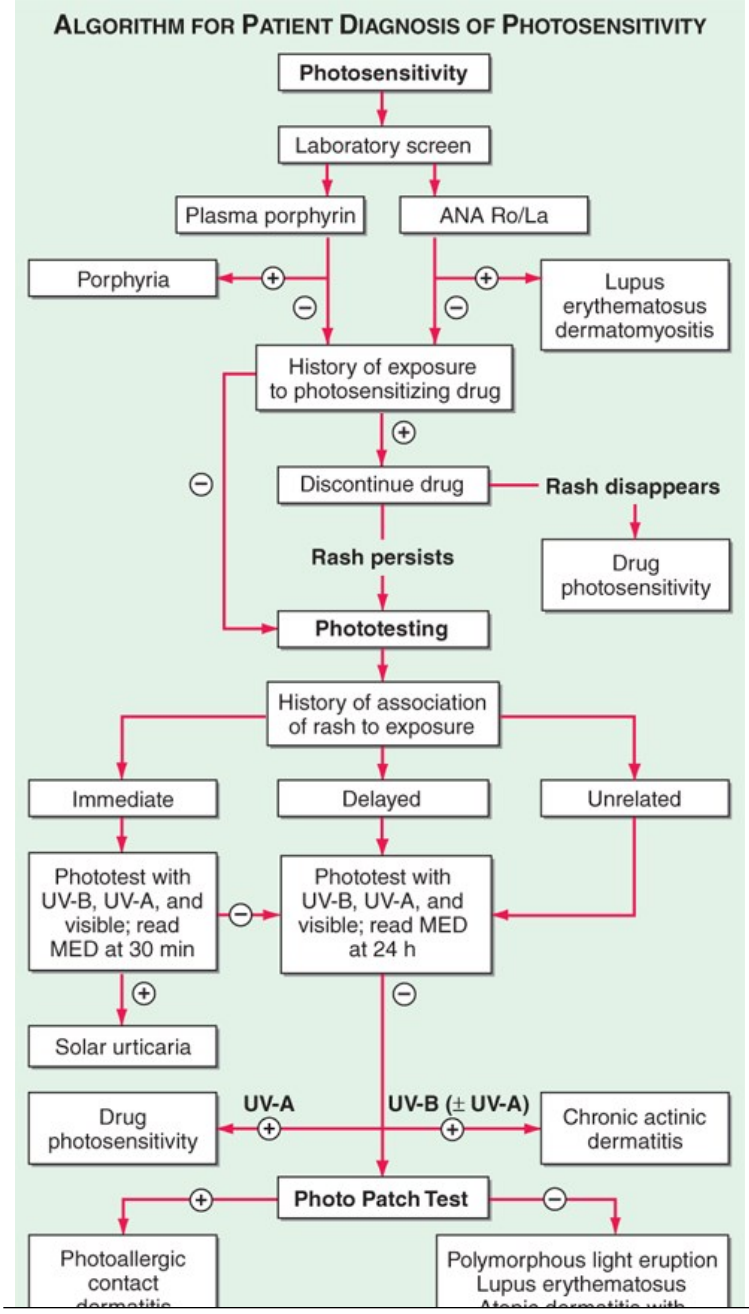
[illegible] patients with skin phototypes IV–VI). Some studies reported that patients with EPP had a moderate increase in tolerance to sunlight after taking oral β-carotene, which may provide this effect by quenching oxygen free radicals.

Rigorous sunlight protection is essential in the management of EPP. Notably, the U.S. Food and Drug Administration (FDA) has approved a synthetic peptide analogue of α-MSH, afamelanotide, in patients with EPP. This drug increases skin pigmentation through melanogenesis, and patients receiving it tolerate sun exposure without pain for longer periods of time and have an improved quality of life as compared to untreated patients. Interestingly, initial studies suggest that afamelanotide may also be beneficial when combined with narrow-band UV-B in the treatment of patients with vitiligo (in patients with skin phototypes IV–VI). Some studies reported that patients with EPP had a moderate increase in tolerance to sunlight after taking oral β-carotene, which may provide this effect by quenching oxygen free radicals.

An algorithm for managing patients with photosensitivity is presented in **Fig. 61-1**.

FIGURE 61-1

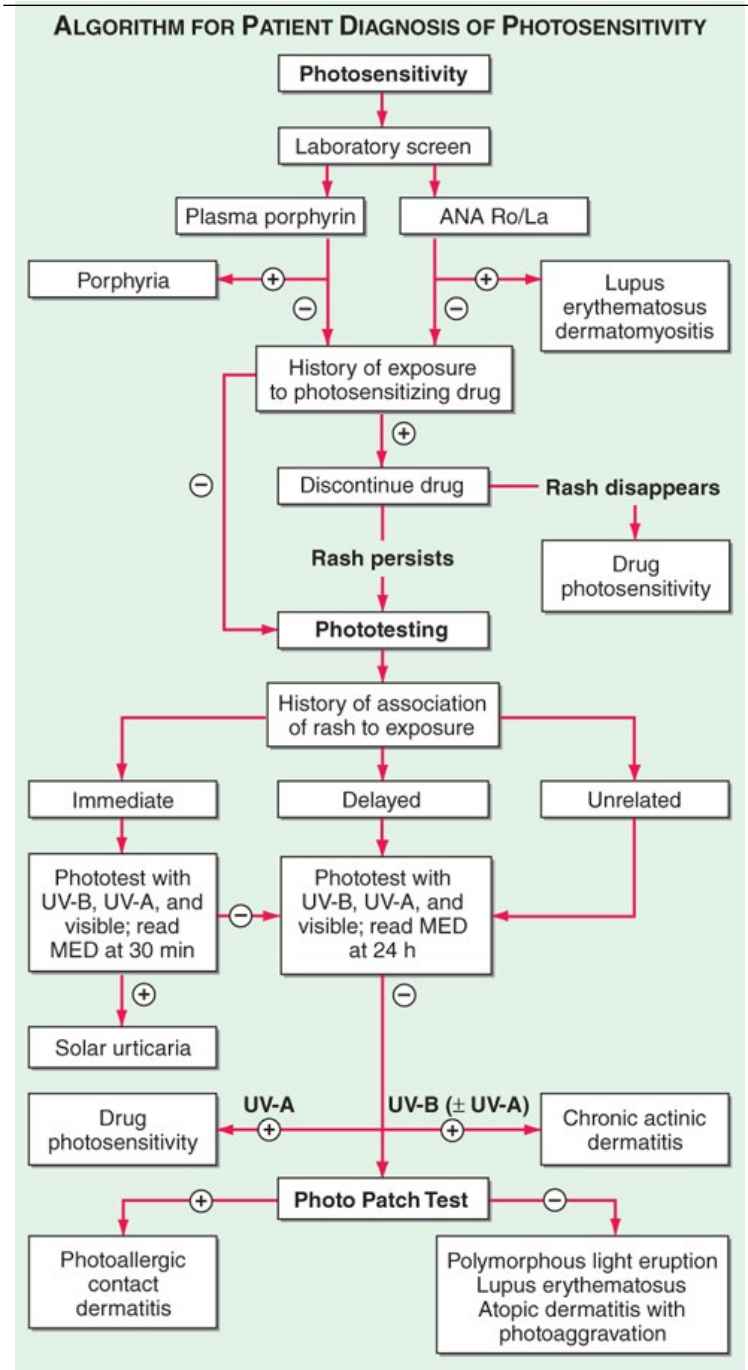
Algorithm for the diagnosis of a patient with photosensitivity. ANA, antinuclear antibody; MED, minimal erythemal dose; UV-A and UV-B, ultraviolet spectrum segments including wavelengths of 320–400 nm and 290–320 nm, respectively.



Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility
Copyright © McGraw Hill. All rights reserved.

PHOTOPROTECTION

ultraviolet spectrum segments including wavelengths of 320–400 nm and 290–320 nm, respectively.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

PHOTOPROTECTION

Since photosensitivity of the skin results from exposure to sunlight, it follows that absolute avoidance of sunlight will eliminate these disorders. However, contemporary lifestyles make this approach impractical for most individuals. Thus, better approaches to photoprotection have been sought. Natural photoprotection is provided by structural proteins in the epidermis, particularly keratins and melanin. The amount of melanin and its distribution in cells are genetically regulated, and individuals of darker complexion (skin types IV–VI) are at decreased risk for the development of acute sunburn and cutaneous malignancy. Other forms of photoprotection include clothing and sunscreens. Clothing constructed of tightly woven sun-protective fabrics, irrespective of color, affords substantial protection. Wide-brimmed hats, long sleeves, and trousers all reduce direct exposure.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

from UV-B but not from UV-A. The FDA monograph stipulates that sunscreens must be rated on a scale ranging from minimal (SPF ≥2 and <12) to moderate (SPF ≥12 and <30) to high (SPF ≥30, labeled as 30+).

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

protective fabrics, irrespective of color, affords substantial protection. Wide-brimmed hats, long sleeves, and trousers all reduce direct exposure.

Sunscreens are now considered over-the-counter drugs, and a monograph from the FDA has recognized category I ingredients as safe and effective. Those ingredients are listed in **Table 61-5**. Sunscreens are rated for their photoprotective effect by their sun protection factor (SPF). The SPF is simply a ratio of the time required to produce sunburn erythema with and without sunscreen application. The SPF of most sunscreens reflects protection from UV-B but not from UV-A. The FDA monograph stipulates that sunscreens must be rated on a scale ranging from minimal (SPF ≥2 and <12) to moderate (SPF ≥12 and <30) to high (SPF ≥30, labeled as 30+).

TABLE 61-5

FDA Category I Monographed Sunscreen Ingredients

INGREDIENTS	MAXIMUM CONCENTRATION, %
p-Aminobenzoic acid (PABA)	15
Avobenzone	3
Cinoxate	3
Dioxybenzone (benzophenone-8)	3
Ecamsule	15
Homosalate	15
Methyl anthranilate	5
Octocrylene	10
Octyl methoxycinnamate	7.5
Octyl salicylate	5
Oxybenzone (benzophenone-3)	6
Padimate O (octyl dimethyl PABA)	8
Phenylbenzimidazole sulfonic acid	4
Sulisobenzone (benzophenone-4)	10
Titanium dioxide	25
Trolamine salicylate	12
Zinc oxide	25

Abbreviation: FDA, U.S. Food and Drug Administration.

Broad-spectrum sunscreens contain both UV-B–absorbing and UV-A–absorbing chemicals (organic filters). These chemicals absorb UVR and transfer the absorbed energy to surrounding cells. Among these sunscreen ingredients, cinnamates, PABA derivatives, and salicylates absorb UV-B. Benzophenones or ecamsule (terephthalylidene dicamphor sulfonic acid) offer protection against UV-B and UV-A2, whereas avobenzone protects mainly against UV-A1. In contrast, physical UV blockers (zinc oxide and titanium dioxide) absorb or reflect UVR and offer broad-spectrum protection.

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

sun protection.

Some degree of photoprotection can be achieved by limiting the time of sun exposure during the day. Since a large part of an individual's total lifetime

Benzophenones or ecamsule (terephthalylidene dicamphor sulfonic acid) offer protection against UV-B and UV-A2, whereas avobenzone protects mainly against UV-A1. In contrast, physical UV blockers (zinc oxide and titanium dioxide) absorb or reflect UVR and offer broad-spectrum protection against UV-B and UV-A. In addition to light absorption, a critical determinant of the sustained photoprotective effect of sunscreens is their water resistance. Sunscreen products with an SPF of 30 or higher, broad-spectrum coverage, and water or sweat resistance are recommended for adequate sun protection.

Some degree of photoprotection can be achieved by limiting the time of sun exposure during the day. Since a large part of an individual's total lifetime sun exposure may occur by age 18, it is important to educate parents and young children about the hazards of sunlight. Eliminating exposure at midday will substantially reduce lifetime UVR exposure.

PHOTOTHERAPY AND PHOTOCHEMOTHERAPY

UVR can be used therapeutically. The administration of UV-B alone or in combination with topically applied agents can induce remissions of many dermatologic diseases, including psoriasis, atopic dermatitis, and vitiligo. In particular, narrow-band UV-B treatments (with fluorescent bulbs emitting radiation at ~311 nm) have enhanced efficacy over that obtained with broad-band UV-B in the treatment of psoriasis.

Photochemotherapy in which topically applied or systemically administered psoralens are combined with UV-A (PUVA) is effective in treating psoriasis and the early stages of cutaneous T-cell lymphoma and vitiligo. Psoralens are tricyclic furocoumarins that, when intercalated into DNA and exposed to UV-A, form adducts with pyrimidine bases and eventually form DNA cross-links. These structural changes are thought to decrease DNA synthesis and to be related to the amelioration of psoriasis. Why PUVA photochemotherapy is effective in cutaneous T-cell lymphoma is only partially understood, but it has been shown to induce apoptosis of atypical T lymphocyte populations in the skin. Consequently, direct treatment of circulating atypical lymphocytes by extracorporeal photochemotherapy (photopheresis) has been used in Sézary syndrome as well as in other severe systemic diseases with circulating atypical lymphocytes, such as graft-versus-host disease.

In addition to its effects on DNA, PUVA photochemotherapy stimulates epidermal thickening and melanin synthesis; the latter property, together with its anti-inflammatory effects, provides the rationale for use of PUVA in the depigmenting disease vitiligo. Oral 8-methoxypsoralen and UV-A appear to be most effective in this regard, but as many as 100 treatments extending over 12–18 months may be required for satisfactory repigmentation.

Not surprisingly, the major side effects of long-term UV-B phototherapy and PUVA photochemotherapy mimic those seen in individuals with chronic sun exposure. Despite these risks, the therapeutic index of these modalities continues to be excellent. It is important to choose the most appropriate phototherapeutic approach for a specific dermatologic disease. For example, narrow-band UV-B has been reported in several studies to be as effective as PUVA photochemotherapy in the treatment of psoriasis but to pose a lower risk of skin cancer development than PUVA.

FURTHER READING

Bernard JJ et al: Photoimmunology: How ultraviolet radiation affects the immune system. Nat Rev Immunol 11:688, 2019.

Fell GL et al: Skin beta-endorphin mediates addiction to UV light. Cell 157:1527, 2014. [PubMed: 24949966]

Harms PW et al: The biology and treatment of Merkel cell carcinoma: Current understanding and research priorities. Nat Rev Clin Oncol 15:763, 2018. [PubMed: 30287935]

Jansen R et al: Photoprotection: Part II. Sunscreen: Development, efficacy, and controversies. J Am Acad Dermatol 69:867, 2013. [PubMed: 24238180]

Lo JA et al: The melanoma revolution: From UV carcinogenesis to a new era in therapeutics. Science 346:945, 2014. [PubMed: 25414302]

Martincorena I et al: Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science 348:880, 2015. [PubMed: 25999502]

Sanchez-Danes A et al: Defining the clonal dynamics leading to mouse skin tumour initiation. Nature 536:298, 2016. [PubMed: 27459053]

Downloaded 2022-1-3 5:50 P Your IP is 141.216.78.40
Chapter 61: Photosensitivity and Other Reactions to Sunlight, Alexander G. Marneros; David R. Bickers
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 65: Bleeding and Thrombosis

Barbara A. Konkle

INTRODUCTION

The human hemostatic system provides a natural balance between procoagulant and anticoagulant forces. The procoagulant forces include platelet adhesion and aggregation and fibrin clot formation; anticoagulant forces include the natural inhibitors of coagulation and fibrinolysis. Under normal circumstances, hemostasis is regulated to promote blood flow; however, it is also prepared to clot blood rapidly to arrest blood flow and prevent exsanguination. After bleeding is successfully halted, the system remodels the damaged vessel to restore normal blood flow. The major components of the hemostatic system, which function in concert, are (1) platelets and other formed elements of blood, such as monocytes and red cells; (2) plasma proteins (the coagulation and fibrinolytic factors and inhibitors); and (3) the vessel wall.

STEPS OF NORMAL HEMOSTASIS

PLATELET PLUG FORMATION

On vascular injury, platelets adhere to the site of injury, usually the denuded vascular intimal surface. Platelet adhesion is mediated primarily by von Willebrand factor (VWF), a large multimeric protein present in both plasma and the extracellular matrix of the subendothelial vessel wall, which serves as the primary "molecular glue," providing sufficient strength to withstand the high levels of shear stress that would tend to detach them with the flow of blood. Platelet adhesion is also facilitated by direct binding to subendothelial collagen through specific platelet membrane collagen receptors.

Platelet adhesion results in subsequent platelet activation and aggregation. This process is enhanced and amplified by humoral mediators in plasma (e.g., epinephrine, thrombin); mediators released from activated platelets (e.g., adenosine diphosphate, serotonin); and vessel wall extracellular matrix constituents that come in contact with adherent platelets (e.g., collagen, VWF). Activated platelets undergo the release reaction, during which they secrete contents that further promote aggregation and inhibit the naturally anticoagulant endothelial cell factors. During platelet aggregation (platelet-platelet interaction), additional platelets are recruited from the circulation to the site of vascular injury, leading to the formation of an occlusive platelet thrombus. The platelet plug is anchored and stabilized by the developing fibrin mesh.

The platelet glycoprotein (Gp) IIb/IIIa ($\alpha_{IIb}\beta_3$) complex is the most abundant receptor on the platelet surface. Platelet activation converts the normally inactive Gp IIb/IIIa receptor into an active receptor, enabling binding to fibrinogen and VWF. Because the surface of each platelet has about 50,000 Gp IIb/IIIa–binding sites, numerous activated platelets recruited to the site of vascular injury can rapidly form an occlusive aggregate by means of a dense network of intercellular fibrinogen bridges.

FIBRIN CLOT FORMATION

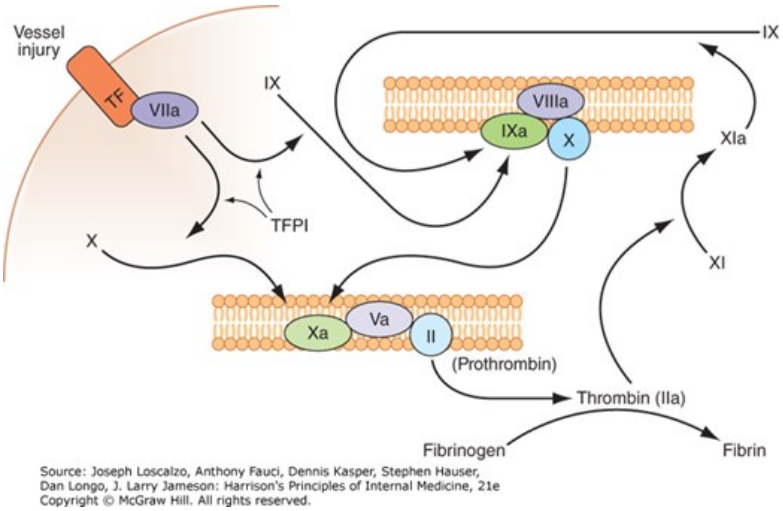
Plasma coagulation proteins (*clotting factors*) normally circulate in plasma in their inactive forms. The sequence of coagulation protein reactions that culminate in the formation of fibrin was originally described as a *waterfall* or a *cascade*. Two pathways of blood coagulation have been described in the past: the so-called extrinsic, or tissue factor, pathway and the so-called intrinsic, or contact activation, pathway. We now know that coagulation is normally initiated through tissue factor (TF) exposure and activation through the classic *extrinsic pathway* but with critically important amplification through elements of the classic *intrinsic pathway*, as illustrated in **Fig. 65-1**. These reactions take place on phospholipid surfaces, usually the activated platelet surface. Coagulation testing in the laboratory can reflect other influences due to the artificial nature of the in vitro systems used (see below).

FIGURE 65-1

Coagulation initiated by tissue factor (TF) [illegible] making coagulation dependent on the amplification loop through FIX/FVIII. Coagulation requires calcium (not shown) and takes place on phospholipid surfaces, usually the activated platelet membrane.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Coagulation is initiated by tissue factor (TF) exposure, which, with factor (F) VIIa, activates FIX and FX, which in turn, with FVIII and FV as cofactors, respectively, results in thrombin formation and subsequent conversion of fibrinogen to fibrin. Thrombin activates FXI, FVIII, and FV, amplifying the coagulation signal. Once the TF/FVIIa/FXa complex is formed, tissue factor pathway inhibitor (TFPI) inhibits the TF/FVIIa pathway, making coagulation dependent on the amplification loop through FIX/FVIII. Coagulation requires calcium (not shown) and takes place on phospholipid surfaces, usually the activated platelet membrane.



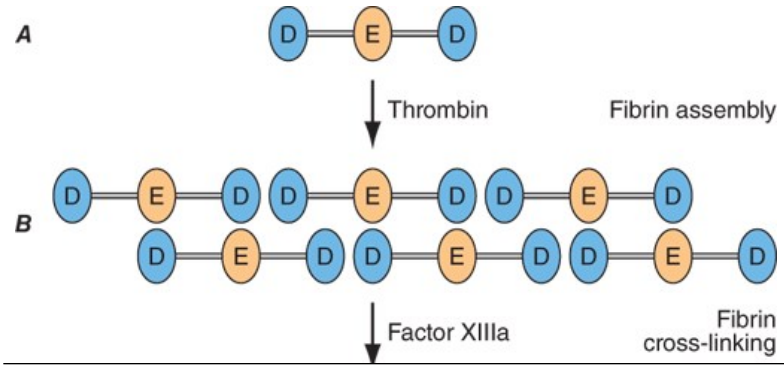
Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The immediate trigger for coagulation is vascular damage that exposes blood to TF that is constitutively expressed on the surfaces of subendothelial cellular components of the vessel wall, such as smooth muscle cells and fibroblasts. TF is also present in circulating microparticles, presumably shed from cells including monocytes and platelets. TF binds the serine protease factor VIIa; the complex activates factor X to factor Xa. Alternatively, the complex can indirectly activate factor X by initially converting factor IX to factor IXa, which then activates factor X. The participation of factor XI in hemostasis is not dependent on its activation by factor XIIa but rather on its positive feedback activation by thrombin. Thus, factor XIa functions in the propagation and amplification, rather than in the initiation, of the coagulation cascade. The role of factor XIIa in activation of factor XI is not fully elucidated, but studies suggest it may be a mechanism to promote thrombosis.

Factor Xa can be formed through the actions of either the TF/factor VIIa complex or factor IXa (with factor VIIIa as a cofactor) and converts prothrombin to thrombin, the pivotal protease of the coagulation system. The essential cofactor for this reaction is factor Va. Like the homologous factor VIIIa, factor Va is produced by thrombin-induced limited proteolysis of factor V. Thrombin is a multifunctional enzyme that converts soluble plasma fibrinogen to an insoluble fibrin matrix. Fibrin polymerization involves an orderly process of intermolecular associations **(Fig. 65-2)**. Thrombin also activates factor XIII (fibrin-stabilizing factor) to factor XIIIa, which covalently cross-links and thereby stabilizes the fibrin clot.

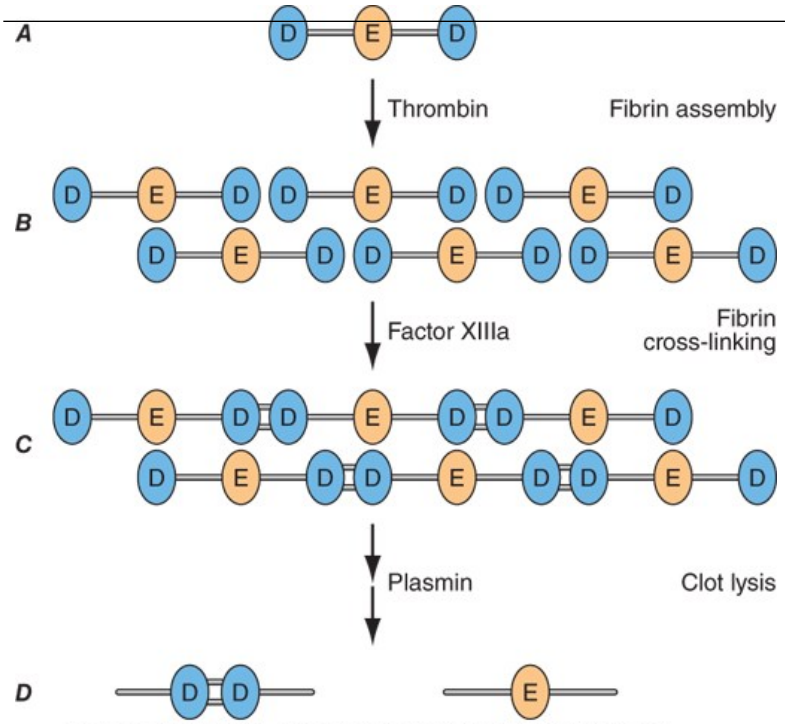
FIGURE 65-2

Fibrin formation and dissolution. (***A***) Fibrinogen is a trinodular structure consisting of two D domains and one E domain. Thrombin activation results in an ordered lateral assembly of protofibrils (***B***) with noncovalent associations. Factor XIIIa cross-links the D domains on adjacent molecules (***C***). Fibrin and fibrinogen (not shown) lysis by plasmin occurs at discrete sites and results in intermediary fibrin(ogen) degradation products (not shown). D-Dimers are the product of complete lysis of fibrin (***D***), maintaining the cross-linked D domains.



Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Fibrin formation and dissolution. (*A*) Fibrinogen is a trinodular structure consisting of two D domains and one E domain. Thrombin activation



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The assembly of the clotting factors on activated cell membrane surfaces greatly accelerates their reaction rates and also serves to localize blood clotting to sites of vascular injury. The critical cell membrane components, acidic phospholipids, are not normally exposed on resting cell membrane surfaces. However, when platelets, monocytes, and endothelial cells are activated by vascular injury or inflammatory stimuli, the procoagulant head groups of the membrane anionic phospholipids become translocated to the surfaces of these cells or released as part of microparticles, making them available to support and promote the plasma coagulation reactions.

ANTITHROMBOTIC MECHANISMS

Several physiologic antithrombotic mechanisms act in concert to prevent clotting under normal circumstances. These mechanisms operate to preserve blood fluidity and to limit blood clotting to specific focal sites of vascular injury. Endothelial cells have many antithrombotic effects. They produce prostacyclin, nitric oxide, and ectoADPase/CD39, which act to inhibit platelet binding, secretion, and aggregation. Endothelial cells produce anticoagulant factors including heparan proteoglycans, TF pathway inhibitor, and thrombomodulin. They also activate fibrinolytic mechanisms through the production of tissue plasminogen activator, urokinase, plasminogen activator inhibitors, and annexin-2.

Antithrombin is the major plasma protease inhibitor of thrombin and the other clotting factors in coagulation. Antithrombin neutralizes thrombin and other activated coagulation factors by forming a complex between the active site of the enzyme and the reactive center of antithrombin. The rate of formation of these inactivating complexes increases by a factor of several thousand in the presence of heparin. Antithrombin inactivation of thrombin and other activated clotting factors occurs physiologically on vascular surfaces, where glycosaminoglycans, including heparan sulfates, are present to catalyze these reactions. Inherited quantitative or qualitative deficiencies of antithrombin lead to a lifelong predisposition to venous thromboembolism.

Protein C is a plasma glycoprotein that becomes an anticoagulant when it is activated by thrombin. The thrombin-induced activation of protein C occurs physiologically on thrombomodulin, a transmembrane proteoglycan-binding site for thrombin on endothelial cell surfaces. The binding of protein C to its receptor on endothelial cells places it in proximity to the thrombin-thrombomodulin complex, thereby enhancing its activation efficiency. (See **Fig. 65-3**.) Activated protein C acts as an anticoagulant by cleaving and inactivating activated factors V and VIII. This reaction is accelerated by a cofactor, protein S, which, like protein C, is a glycoprotein that undergoes vitamin K–dependent posttranslational modification. Quantitative or qualitative deficiencies of protein C or protein S, or resistance to the action of activated protein C by a specific variant at its target cleavage site in factor Va (factor V Leiden), lead to hypercoagulable states.

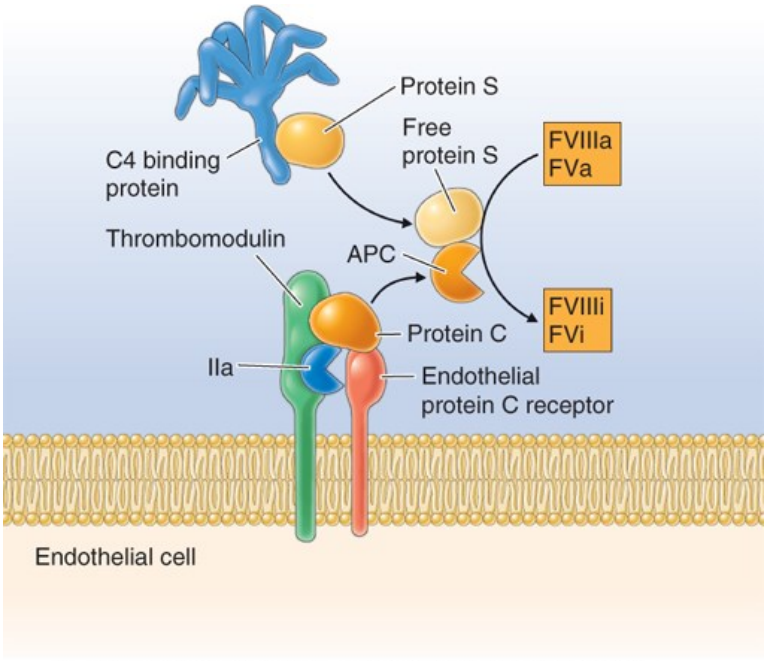
Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

thrombomodulin and protein C bound to the endothelial protein C receptor (EPCR). Activated protein C (APC) with free protein S converts activated factors (F) VIII and V to inactivate forms, thus in turn decreasing thrombin generation. C4BP, C4 binding protein; EC, endothelial cell; F, factor; IIa,

cleavage site in factor Va (factor V Leiden), lead to hypercoagulable states.

FIGURE 65-3

The activated protein C pathway in regulation of thrombosis. Thrombin generation results in protein C activation through interaction with thrombomodulin and protein C bound to the endothelial protein C receptor (EPCR). Activated protein C (APC) with free protein S converts activated factors (F) VIII and V to inactivate forms, thus in turn decreasing thrombin generation. C4BP, C4 binding protein; EC, endothelial cell; F, factor; IIa, thrombin; PC, protein C; PS, protein S; TM, thrombomodulin.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

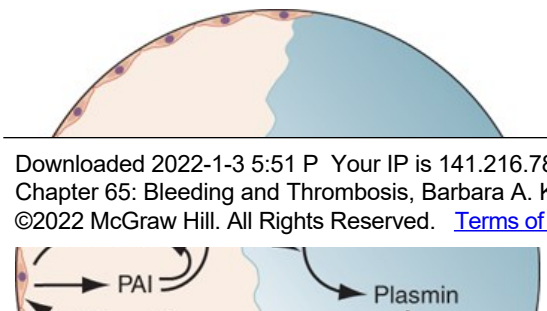
Tissue factor pathway inhibitor (TFPI) is a plasma protease inhibitor that regulates the TF-induced extrinsic pathway of coagulation. TFPI inhibits the TF/factor VIIa/factor Xa complex, essentially turning off the TF/factor VIIa initiation of coagulation, which then becomes dependent on the "amplification loop" via factor XI and factor VIII activation by thrombin. TFPI is bound to lipoprotein and can also be released by heparin from endothelial cells, where it is bound to glycosaminoglycans, and from platelets. The heparin-mediated release of TFPI may play a role in the anticoagulant effects of unfractionated and low-molecular-weight heparins.

THE FIBRINOLYTIC SYSTEM

Any thrombin that escapes the inhibitory effects of the physiologic anticoagulant systems is available to convert fibrinogen to fibrin. In response, the endogenous fibrinolytic system is then activated to dispose of intravascular fibrin and thereby maintain or reestablish the patency of the circulation. Just as thrombin is the key protease enzyme of the coagulation system, plasmin is the major protease enzyme of the fibrinolytic system, acting to digest fibrin to fibrin degradation products. The general scheme of fibrinolysis and its control is shown in **Fig. 65-4**.

FIGURE 65-4

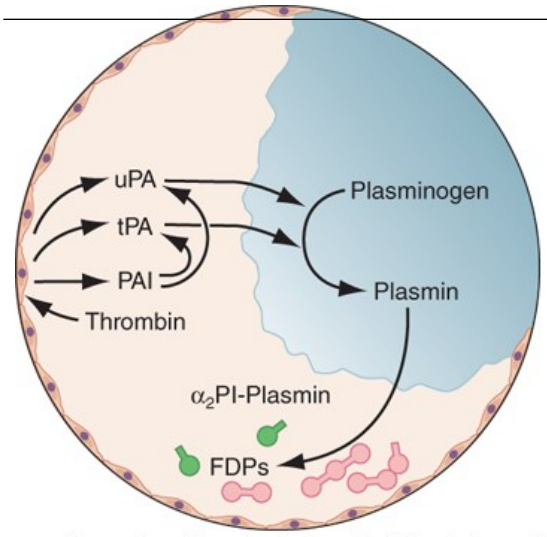
A schematic diagram of the fibrinolytic system. Tissue plasminogen activator (tPA) is released from endothelial cells, binds the fibrin clot, and activates plasminogen to plasmin. Release of plasminogen activator inhibitors (PAI-1 and PAI-2) inhibits tPA and urokinase (uPA). Excess fibrin is degraded by plasmin to distinct degradation products [FDPs (D-dimers)]. Any free plasmin is complexed with α_2-antiplasmin (α_2PI). PAI, plasminogen activator inhibitor; uPA, urokinase-type plasminogen activator.



Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The plasminogen activators, tissue type plasminogen activator (tPA) and the urokinase-type plasminogen activator (uPA), cleave the Arg560-Val561 bond of plasminogen to generate the active enzyme plasmin. The lysine-binding sites of plasmin (and plasminogen) permit it to bind to fibrin, so that physiologic fibrinolysis is "fibrin specific." Both plasminogen (through its lysine-binding sites) and tPA possess specific affinity for fibrin and thereby bind selectively to clots. The assembly of a ternary complex, consisting of fibrin, plasminogen, and tPA, promotes the localized interaction between plasminogen and tPA and greatly accelerates the rate of plasminogen activation to plasmin. Moreover, partial degradation of fibrin by plasmin exposes new plasminogen and tPA-binding sites in carboxy-terminus lysine residues of fibrin fragments to enhance these reactions further. This creates a highly efficient mechanism to generate plasmin focally on the fibrin clot, which then becomes plasmin's substrate for digestion to fibrin degradation products.

Plasmin cleaves fibrin at distinct sites of the fibrin molecule, leading to the generation of characteristic fibrin fragments during the process of fibrinolysis (Fig. 65-2). The sites of plasmin cleavage of fibrin are the same as those in fibrinogen. However, when plasmin acts on covalently cross-linked fibrin, D-dimers are released; hence, D-dimers can be measured in plasma as a relatively specific test of fibrin (rather than fibrinogen) degradation. D-Dimer assays can be used as sensitive markers of blood clot formation and have been validated for clinical use to exclude the diagnosis of deep venous thrombosis (DVT) and pulmonary embolism in selected populations. D-Dimer levels increase with age. A higher cut-off value to rule out venous thromboembolism (VTE) in the elderly has been proposed but is controversial.

Physiologic regulation of fibrinolysis occurs primarily at three levels: (1) plasminogen activator inhibitors (PAIs), specifically PAI-1 and PAI-2, inhibit the physiologic plasminogen activators; (2) the thrombin-activatable fibrinolysis inhibitor (TAFI) limits fibrinolysis; and (3) α_2-antiplasmin inhibits plasmin. PAI-1 is the primary inhibitor of tPA and uPA in plasma. TAFI cleaves the N-terminal lysine residues of fibrin, which aid in localization of plasmin activity. α_2-Antiplasmin is the main inhibitor of plasmin in human plasma, inactivating any nonfibrin clot–associated plasmin.

APPROACH TO THE PATIENT

Bleeding and Thrombosis

CLINICAL PRESENTATION

Disorders of hemostasis may be either inherited or acquired. A detailed personal and family history is key in determining the chronicity of symptoms and the likelihood of the disorder being inherited, as well as providing clues to underlying conditions that have contributed to the bleeding or thrombotic state. In addition, the history can give clues as to the etiology by determining (1) the bleeding (mucosal and/or joint) or thrombosis (arterial and/or venous) site and (2) whether an underlying bleeding or clotting tendency was enhanced by another medical condition or the introduction of medications or dietary supplements.

History of Bleeding

A history of bleeding is the most important predictor of bleeding risk. In evaluating a patient for a bleeding disorder, a history of at-risk situations

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

of primary hemostasis or platelet plug formation. Disorders affecting primary hemostasis are shown in **Table 65-1**.

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

History of Bleeding

A history of bleeding is the most important predictor of bleeding risk. In evaluating a patient for a bleeding disorder, a history of at-risk situations, including the response to past surgeries, should be assessed. Does the patient have a history of spontaneous or trauma/surgery-induced bleeding? Spontaneous hemarthroses are a hallmark of moderate and severe factor VIII and IX deficiency and, in rare circumstances, of other clotting factor deficiencies. Mucosal bleeding symptoms are more suggestive of underlying platelet disorders or von Willebrand disease (VWD), termed *disorders of primary hemostasis or platelet plug formation.* Disorders affecting primary hemostasis are shown in **Table 65-1**.

A bleeding score has been validated as a tool to predict patients more likely to have type 1 VWD (International Society on Thrombosis and Haemostasis Bleeding Assessment Tool [*www.isth.org/resource/resmgr/ssc/isth-ssc_bleeding_assessment.pdf*]), and a self-administered form has been validated. This is the most useful tool in excluding the diagnosis of a bleeding disorder, thus avoiding unnecessary testing, and is recommended by 2021 guidelines for screening for VWD in primary care. Bleeding symptoms that are more common in patients with bleeding disorders include prolonged bleeding with surgery, dental procedures and extractions, and/or trauma; heavy menstrual bleeding or postpartum hemorrhage; and large bruises (often described with lumps).

Easy bruising and heavy menstrual bleeding are common complaints in patients with and without bleeding disorders. Easy bruising can also be a sign of medical conditions in which there is no identifiable coagulopathy; instead, the conditions are caused by an abnormality of blood vessels or their supporting connective tissues. In Ehlers-Danlos syndrome, there may be posttraumatic bleeding and a history of joint hyperextensibility. Cushing's syndrome, chronic steroid use, and aging result in changes in skin and subcutaneous tissue, and subcutaneous bleeding occurs in response to minor trauma. The latter has been termed *senile purpura*.

Epistaxis is a common symptom, particularly in children and in dry climates, and may not reflect an underlying bleeding disorder. However, it is the most common symptom in hereditary hemorrhagic telangiectasia and in boys with VWD. Clues that epistaxis is a symptom of an underlying bleeding disorder include lack of seasonal variation and bleeding that requires medical evaluation or treatment, including cauterization. Bleeding with eruption of primary teeth is seen in children with more severe bleeding disorders, such as moderate and severe hemophilia. It is uncommon in children with mild bleeding disorders. Patients with disorders of primary hemostasis (platelet adhesion) may have increased bleeding after dental cleanings and other procedures that involve gum manipulation.

Heavy menstrual bleeding is defined quantitatively as a loss of >80 mL of blood per cycle, based on the quantity of blood loss required to produce iron-deficiency anemia. A complaint of heavy menses is subjective and has a poor correlation with excessive blood loss. Predictors of heavy menstrual bleeding include bleeding resulting in iron-deficiency anemia or a need for blood transfusion, passage of clots >1 inch in diameter, and changing a pad or tampon more than hourly. Heavy menstrual bleeding is a common symptom in women with underlying bleeding disorders and is reported in the majority of women with VWD, factor XI deficiency, platelet function disorders, and hemophilia, including genetic carriers with borderline-normal factor levels. Women with underlying bleeding disorders are more likely to have other bleeding symptoms, including bleeding after dental extractions and postoperative and postpartum bleeding, and are much more likely to have heavy menstrual bleeding beginning at menarche than women with heavy menstrual bleeding due to other causes. Heavy menstrual bleeding may results in iron deficiency and is documented to have significant adverse effects on quality of life.

Postpartum hemorrhage is a common symptom in women with underlying bleeding disorders. In women with type 1 VWD or hemophilia A in whom levels of VWF and factor VIII usually normalize during pregnancy, postpartum hemorrhage may be delayed. Women with a history of postpartum hemorrhage may have a higher risk of recurrence with subsequent pregnancies. Women with underlying bleeding disorders are at risk for other reproductive tract bleeding, including rupture of ovarian cysts with intraabdominal hemorrhage.

Tonsillectomy is a major hemostatic challenge, because intact hemostatic mechanisms are essential to prevent excessive bleeding from the tonsillar bed. Bleeding may occur early after surgery or after approximately 7 days postoperatively, with loss of the eschar at the operative site. Similar delayed bleeding is seen after colonic polyp resection. Gastrointestinal (GI) bleeding and hematuria are usually due to underlying pathology, and procedures to identify and treat the bleeding site should be undertaken, even in patients with known bleeding disorders. VWD, particularly types 2 and 3, is associated with angiodysplasia of the bowel and GI bleeding.

Hemarthroses and spontaneous muscle hematomas are characteristic of moderate or severe congenital factor VIII or IX deficiency. They can also be seen in moderate and severe deficiencies of fibrinogen, prothrombin, and factors V, VII, and X. Spontaneous hemarthroses occur rarely in other bleeding disorders except for severe VWD, with associated factor VIII levels <5%. Muscle and soft tissue bleeds are also common in acquired factor VIII deficiency. Bleeding into a joint results in severe pain and swelling, as well as loss of function, but is rarely associated with discoloration from bruising around the joint. Life-threatening sites of bleeding include bleeding into the oropharynx, where bleeding can obstruct the airway, into the

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Prohemorrhagic Effects of Medications and Dietary Supplements

Aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) that inhibit cyclooxygenase 1 impair primary hemostasis and may exacerbate

VIII deficiency. Bleeding into a joint results in severe pain and swelling, as well as loss of function, but is rarely associated with discoloration from bruising around the joint. Life-threatening sites of bleeding include bleeding into the oropharynx, where bleeding can obstruct the airway, into the central nervous system, and into the retroperitoneum. Central nervous system bleeding is the major cause of bleeding-related deaths in patients with severe congenital factor deficiencies.

Prohemorrhagic Effects of Medications and Dietary Supplements

Aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) that inhibit cyclooxygenase 1 impair primary hemostasis and may exacerbate bleeding from another cause or even unmask a previously occult mild bleeding disorder such as VWD. All NSAIDs, however, can precipitate GI bleeding, which may be more severe in patients with underlying bleeding disorders. The aspirin effect on platelet function lasts for the life of the platelet; however, in individuals with typical platelet turnover, the functional defect reverts to near-normal within 2–3 days after the last dose. The effect of other NSAIDs is shorter, as the inhibitor effect is reversed when the drug is removed. Inhibitors of the ADP $P2Y_{12}$ receptor (clopidogrel, prasugrel, and ticagrelor) inhibit ADP-mediated platelet aggregation and, like NSAIDs, can precipitate or exacerbate bleeding symptoms. The risk of bleeding with these drugs is higher than with NSAIDs.

Many herbal supplements can impair hemostatic function **(Table 65-2)**. Some are more convincingly associated with a bleeding risk than others. Fish oil or concentrated omega-3 fatty acid supplements impair platelet function. They alter platelet biochemistry to produce more PGI_3, a more potent platelet inhibitor than prostacyclin (PGI_2), and more thromboxane A_3, a less potent platelet activator than thromboxane A_2. In fact, diets naturally rich in omega-3 fatty acids can result in a prolonged bleeding time and abnormal platelet aggregation studies, but the actual associated bleeding risk is unclear. Vitamin E appears to inhibit protein kinase C–mediated platelet aggregation and nitric oxide production. In patients with unexplained bruising or bleeding, it is prudent to review any new medications or supplements and discontinue those that may be associated with bleeding.

Underlying Systemic Diseases That Cause or Exacerbate a Bleeding Tendency

Acquired bleeding disorders are commonly secondary to, or associated with, systemic disease. The clinical evaluation of a patient with a bleeding tendency must therefore include a thorough assessment for evidence of underlying disease. Bruising or mucosal bleeding may be the presenting complaint in liver disease, severe renal impairment, hypothyroidism, paraproteinemias or amyloidosis, and conditions causing bone marrow failure. All coagulation factors are synthesized in the liver, and hepatic failure results in combined factor deficiencies. This is often compounded by thrombocytopenia and portal hypertension. Coagulation factors II, VII, IX, and X and proteins C, S, and Z are dependent on vitamin K for posttranslational modification. Although vitamin K is required in both procoagulant and anticoagulant processes, the phenotype of vitamin K deficiency or the warfarin effect on coagulation is bleeding.

The normal blood platelet count is 150,000–450,000/μL. Thrombocytopenia results from decreased production, increased destruction, and/or sequestration. Although the bleeding risk varies somewhat by the reason for the thrombocytopenia, bleeding rarely occurs in isolated thrombocytopenia at counts >50,000/μL and usually not until <10,000–20,000/μL. Coexisting coagulopathies, as is seen in liver failure or disseminated coagulation; infection; platelet-inhibitory drugs; and underlying medical conditions can all increase the risk of bleeding in the thrombocytopenic patient. Most procedures can be performed in patients with a platelet count of 50,000/μL or greater.

HISTORY OF THROMBOSIS

The risk of thrombosis, like that of bleeding, is influenced by both genetic and environmental factors. The major risk factor for arterial thrombosis is atherosclerosis, whereas for venous thrombosis, the risk factors are immobility, surgery, underlying medical conditions such as malignancy, medications such as hormonal therapy, obesity, and genetic predispositions. Factors that increase risks for venous and for both venous and arterial thromboses are shown in **Table 65-3**.

The most important point in a history related to venous thrombosis is determining whether the thrombotic event was idiopathic (meaning there was no clear precipitating factor) or was a precipitated event. In patients without underlying malignancy, having an idiopathic event is the strongest predictor of recurrence of VTE. In patients who have a vague history of thrombosis, a history of being treated with warfarin or other anticoagulants suggests a past DVT. Age is an important risk factor for venous thrombosis—the risk of DVT increases per decade, with an approximate incidence of 1/100,000 per year in early childhood to 1/200 per year among octogenarians. Family history is helpful in determining if there is a genetic predisposition and how strong that predisposition appears to be. A genetic thrombophilia that confers a relatively small increased risk, such as being a heterozygote for the prothrombin G20210A or factor V Leiden mutation, is a minor determinant of risk in an elderly individual undergoing a high-risk surgical procedure. As illustrated in **Fig. 65-5**, a thrombotic event usually has more than one contributing factor. Predisposing factors must be carefully assessed to determine the risk of recurrent thrombosis and, with consideration of the patient's bleeding risk, determine the

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

LABORATORY EVALUATION

Careful history taking and clinical examination are essential components in the assessment of bleeding and thrombotic risk. The use of laboratory

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

suggests a past DVT. Age is an important risk factor for venous thrombosis—the risk of DVT increases per decade, with an approximate incidence of high-risk surgical procedure. As illustrated in **Fig. 65-5**, a thrombotic event usually has more than one contributing factor. Predisposing factors must be carefully assessed to determine the risk of recurrent thrombosis and, with consideration of the patient's bleeding risk, determine the length of anticoagulation. Testing for inherited thrombophilias in adults should be limited to instances where results would change clinical care. Such instances are rare.

LABORATORY EVALUATION

Careful history taking and clinical examination are essential components in the assessment of bleeding and thrombotic risk. The use of laboratory tests of coagulation complements, but cannot substitute for, clinical assessment. No test exists that provides a global assessment of hemostasis. The bleeding time has been used to assess bleeding risk; however, it does not predict bleeding risk with surgery, and it is not recommended for this indication. The PFA-100, an instrument that measures platelet-dependent coagulation under flow conditions, is more sensitive and specific for VWD than the bleeding time; however, it is not sensitive enough to rule out mild bleeding disorders. PFA-100 closure times are prolonged in patients with some, but not all, inherited platelet disorders. Also, its utility in predicting bleeding risk has not been determined. Thromboelastography can be useful in guiding intraoperative transfusion and is being explored in other settings, but is not broadly applicable for the diagnosis of disorders of hemostasis and thrombosis.

For routine preoperative and preprocedure testing, an abnormal prothrombin time (PT) may detect liver disease or vitamin K deficiency that had not been previously appreciated. Studies have not confirmed the usefulness of an activated partial thromboplastin time (aPTT) in preoperative evaluations in patients with a negative bleeding history. The primary use of coagulation testing should be to confirm the presence and type of bleeding disorder in a patient with a suspicious clinical history.

Because of the nature of coagulation assays, proper sample acquisition and handling is critical to obtaining valid results. In patients with abnormal coagulation assays who have no bleeding history, repeat studies with attention to these factors frequently results in normal values. Most coagulation assays are performed in sodium citrate anticoagulated plasma that is recalcified for the assay. Because the anticoagulant is in liquid solution and needs to be added to blood in proportion to the plasma volume, incorrectly filled or inadequately mixed blood collection tubes will give erroneous results. These vacutainer tubes should be filled to >90% of the recommended fill, which is usually denoted by a line on the tube. An elevated hematocrit (>55%) can result in a false value due to a decreased plasma-to-anticoagulant ratio.

Screening Assays

The most commonly used screening tests are the PT, aPTT, and platelet count. The PT assesses the factors I (fibrinogen), II (prothrombin), V, VII, and X **(Fig. 65-6)**. The PT measures the time for clot formation of the citrated plasma after recalcification and addition of thromboplastin, a mixture of TF and phospholipids. The sensitivity of the assay varies by the source of thromboplastin. The relationship between defects in secondary hemostasis (fibrin formation) and coagulation test abnormalities is shown in **Table 65-4**. To adjust for this variability, the overall sensitivity of different thromboplastins to reduction of the vitamin K–dependent clotting factors II, VII, IX, and X in anticoagulation patients is expressed as the International Sensitivity Index (ISI). The international normalized ratio (INR) is determined based on the formula: $INR = (PT_{patient}/PT_{normal\ mean})^{ISI}$.

The INR was developed to assess stable anticoagulation due to reduction of vitamin K–dependent coagulation factors; it is commonly used in the evaluation of patients with liver disease. Although it does allow comparison between laboratories, reagent sensitivity as used to determine the ISI is not the same in liver disease as with warfarin anticoagulation. In addition, progressive liver failure is associated with variable changes in coagulation factors; the degree of prolongation of either the PT or the INR only roughly predicts the bleeding risk. Thrombin generation has been shown to be normal in many patients with mild to moderate liver dysfunction. Because the PT only measures one aspect of hemostasis affected by liver dysfunction, we likely overestimate the bleeding risk of a mildly elevated INR in this setting. PT reagents have variable sensitivity to the direct Xa inhibitors, and the PT is usually normal in patients on apixaban.

The aPTT assesses the intrinsic and common coagulation pathways; factors XI, IX, VIII, X, V, and II; fibrinogen; prekallikrein; high-molecular-weight kininogen; and factor XII (Fig. 65-6). The aPTT reagent contains phospholipids derived from either animal or vegetable sources that function as a platelet substitute in the coagulation pathways and includes an activator of the intrinsic coagulation system, such as nonparticulate ellagic acid or the particulate activators kaolin, celite, or micronized silica.

The phospholipid composition of aPTT reagents varies, which influences the sensitivity of individual reagents to clotting factor deficiencies and to inhibitors such as heparin and lupus anticoagulants. Thus, aPTT results will vary from one laboratory to another, and the normal range in the laboratory where the testing occurs should be used in the interpretation. Local laboratories can relate their aPTT values to the therapeutic heparin anticoagulation by correlating aPTT values with direct measurements of heparin activity (anti-Xa or protamine titration assays) in samples from heparinized patients, although correlation between these assays is often poor. The aPTT reagent will vary in sensitivity to individual factor

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Mixing studies are used to evaluate a prolonged aPTT or, less commonly PT, to distinguish between a factor deficiency and an inhibitor. In this assay, normal plasma and patient plasma are mixed in a 1:1 ratio, and the aPTT or PT is determined immediately and after incubation at 37°C for

anticoagulation by correlating aPTT values with direct measurements of heparin activity (anti-Xa or protamine titration assays) in samples from heparinized patients, although correlation between these assays is often poor. The aPTT reagent will vary in sensitivity to individual factor deficiencies and usually becomes prolonged with individual factor deficiencies of 30–50%.

Mixing Studies

Mixing studies are used to evaluate a prolonged aPTT or, less commonly PT, to distinguish between a factor deficiency and an inhibitor. In this assay, normal plasma and patient plasma are mixed in a 1:1 ratio, and the aPTT or PT is determined immediately and after incubation at 37°C for varying times, typically 30, 60, and/or 120 min. With isolated factor deficiencies, the aPTT will correct with mixing and stay corrected with incubation. With aPTT prolongation due to a lupus anticoagulant, the mixing and incubation will show no correction. In acquired neutralizing factor antibodies, notably an acquired factor VIII inhibitor, the initial assay may or may not correct immediately after mixing but will prolong or remain prolonged with incubation at 37°C. Failure to correct with mixing can also be due to the presence of other inhibitors or interfering substances such as heparin, fibrin split products, and paraproteins.

Specific Factor Assays

Decisions to proceed with specific clotting factor assays will be influenced by the clinical situation and the results of coagulation screening tests. Precise diagnosis and effective management of inherited and acquired coagulation deficiencies necessitate quantitation of the relevant factors. When bleeding is severe, specific assays are urgently required to guide appropriate therapy. Individual factor assays are usually performed as modifications of the mixing study, where the patient's plasma is mixed with plasma deficient in the factor being studied. This will correct all factor deficiencies to >50%, thus making prolongation of clot formation due to a factor deficiency dependent on the factor missing from the added plasma.

Testing for Antiphospholipid Antibodies

Antibodies to phospholipids (cardiolipin) or phospholipid-binding proteins (β_2-microglobulin and others) are detected by enzyme-linked immunosorbent assay (ELISA). When these antibodies interfere with phospholipid-dependent coagulation tests, they are termed *lupus anticoagulants*. The aPTT has variability sensitivity to lupus anticoagulants, depending in part on the aPTT reagents used. An assay using a sensitive reagent has been termed an *LA-PTT*. The dilute Russell viper venom test (dRVVT) is a modification of a standard test with the phospholipid reagent decreased, thus increasing the sensitivity to antibodies that interfere with the phospholipid component. These tests, however, are not specific for lupus anticoagulants, because factor deficiencies or other inhibitors will also result in prolongation. Documentation of a lupus anticoagulant requires not only prolongation of a phospholipid-dependent coagulation test but also lack of correction when mixed with normal plasma and correction with the addition of activated platelet membranes or certain phospholipids (e.g., hexagonal phase).

Other Coagulation Tests

The thrombin time and the reptilase time measure fibrinogen conversion to fibrin and are prolonged when the fibrinogen level is low (usually <80–100 mg/dL) or qualitatively abnormal, as seen in inherited or acquired dysfibrinogenemias, or when fibrin/fibrinogen degradation products interfere. The thrombin time, but not the reptilase time, is prolonged in the presence of heparin. The thrombin time is markedly prolonged in the presence of the direct thrombin inhibitor, dabigatran; a dilute thrombin time can be used to assess drug activity. Measurement of anti–factor Xa plasma inhibitory activity is a test frequently used to assess low-molecular-weight heparin (LMWH) levels, as a direct measurement of unfractionated heparin (UFH) activity, or to assess activity of the direct Xa inhibitors rivaroxaban, apixaban, and edoxaban. Drug in the patient sample inhibits the enzymatic conversion of an Xa-specific chromogenic substrate to colored product by factor Xa. Standard curves are created using multiple concentrations of the specific drug and are used to calculate the concentration of anti-Xa activity in the patient plasma.

Laboratory Testing for Thrombophilia

Laboratory assays to detect thrombophilic states include molecular diagnostics and immunologic and functional assays. These assays vary in their sensitivity and specificity for the condition being tested. Furthermore, acute thrombosis, acute illnesses, inflammatory conditions, pregnancy, and medications affect levels of many coagulation factors and their inhibitors. Antithrombin is decreased by heparin and in the setting of acute thrombosis. Protein C and S levels may be increased in the setting of acute thrombosis and are decreased by warfarin. Antiphospholipid antibodies are frequently transiently positive in acute illness. Testing for genetic thrombophilias should, in general, only be performed when there is a strong family history of thrombosis and results would affect clinical decision-making.

Because thrombophilia evaluations are usually performed to assess the need to extend anticoagulation, testing, if indicated, should be performed in a steady state, remote from the acute event. Functional assays, but not genetic assays, will be affected by anticoagulants including warfarin (for vitamin K–dependent proteins) and thrombin and Xa inhibitors and cannot be interpreted in patients on those drugs. In most instances, when discontinuation of anticoagulation is being considered, drugs can be stopped after the initial 3–6 months of treatment, and testing can be

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The bleeding time was used in the past to assess bleeding risk; however, it has not been found to predict bleeding risk with surgery, and it is not recommended for use for this indication. The PFA-100 and similar instruments that measure platelet-dependent coagulation under flow conditions

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

vitamin K–dependent proteins) and thrombin and Xa inhibitors and cannot be interpreted in patients on those drugs. In most instances, when discontinuation of anticoagulation is being considered, drugs can be stopped after the initial 3–6 months of treatment, and testing can be performed at least 3 weeks later.

Measures of Platelet Function

The bleeding time was used in the past to assess bleeding risk; however, it has not been found to predict bleeding risk with surgery, and it is not recommended for use for this indication. The PFA-100 and similar instruments that measure platelet-dependent coagulation under flow conditions are generally more sensitive and specific for platelet disorders and VWD than the bleeding time; however, data are insufficient to support their use to predict bleeding risk or monitor response to therapy, and they will be normal in some patients with platelet disorders or mild VWD. When they are used in the evaluation of a patient with bleeding symptoms, abnormal results require specific testing, such as VWF assays and/or platelet aggregation studies. Because all of these "screening" assays may miss patients with mild bleeding disorders, further studies are needed to define their role in hemostasis testing.

For classic platelet aggregometry, various agonists are added to the patient's platelet-rich plasma or whole blood, and platelet aggregation is measured. Tests of platelet secretion in response to agonists can also be measured. These remain the gold standard for diagnosis of platelet function disorders. However, they are affected by many factors, including numerous medications, and the association between minor defects in these assays and bleeding risk is not clearly established.

TABLE 65-1

Primary Hemostatic (Platelet Plug) Disorders

Defects of Platelet Adhesion
von Willebrand disease
Bernard-Soulier syndrome (absence or dysfunction of platelet Gp Ib-IX-V)
Defects of Platelet Aggregation
Glanzmann's thrombasthenia (absence or dysfunction of platelet glycoprotein [Gp] IIb/IIIa)
Afibrinogenemia
Defects of Platelet Secretion
Decreased cyclooxygenase activity
Drug-induced (aspirin, nonsteroidal anti-inflammatory agents, thienopyridines)
Inherited
Granule storage pool defects
Inherited
Acquired
Nonspecific inherited secretory defects
Nonspecific drug effects
Defect of Platelet Coagulant Activity

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 65-1

Primary Hemostatic (Platelet Plug) Disorders

Defects of Platelet Adhesion
von Willebrand disease
Bernard-Soulier syndrome (absence or dysfunction of platelet Gp Ib-IX-V)
Defects of Platelet Aggregation
Glanzmann's thrombasthenia (absence or dysfunction of platelet glycoprotein [Gp] IIb/IIIa)
Afibrinogenemia
Defects of Platelet Secretion
Decreased cyclooxygenase activity
Drug-induced (aspirin, nonsteroidal anti-inflammatory agents, thienopyridines)
Inherited
Granule storage pool defects
Inherited
Acquired
Nonspecific inherited secretory defects
Nonspecific drug effects
Uremia
Platelet coating (e.g., paraprotein, penicillin)
Defect of Platelet Coagulant Activity
Scott's syndrome

TABLE 65-2

Herbal Supplements Associated with Increased Bleeding

Herbs with Potential Antiplatelet Activity
Ginkgo (*Ginkgo biloba L.*)
Garlic (*Allium sativum*)
Dong quai (*Angelica sinensis*)

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 65-2

Herbal Supplements Associated with Increased Bleeding

Herbs with Potential Antiplatelet Activity
Ginkgo (*Ginkgo biloba L.*)
Garlic (*Allium sativum*)
Bilberry (*Vaccinium myrtillus*)
Ginger (*Gingiber officinale*)
Dong quai (*Angelica sinensis*)
Feverfew (*Tanacetum parthenium*)
Asian ginseng (*Panax ginseng*)
American ginseng (*Panax quinquefolius*)
Siberian ginseng/eleuthero (*Eleutherococcus senticosus*)
Turmeric (*Circuma longa*)
Meadowsweet (*Filipendula ulmaria*)
Willow (*Salix* spp.)
Coumarin-Containing Herbs
Motherwort (*Leonurus cardiaca*)
Chamomile (*Matricaria recutita, Chamaemelum mobile*)
Horse chestnut (*Aesculus hippocastanum*)
Red clover (*Trifolium pratense*)
Fenugreek (*Trigonella foenum-graecum*)

TABLE 65-3

Some Risk Factors for Thrombosis

VENOUS	VENOUS AND ARTERIAL
Inherited	**Inherited**
Factor V Leiden	Homocystinuria
Prothrombin G20210A	Dysfibrinogenemia
Antithrombin deficiency	**Acquired**
Age	Polycythemia vera
Previous thrombosis	Essential thrombocythemia

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

TABLE 65-3

Some Risk Factors for Thrombosis

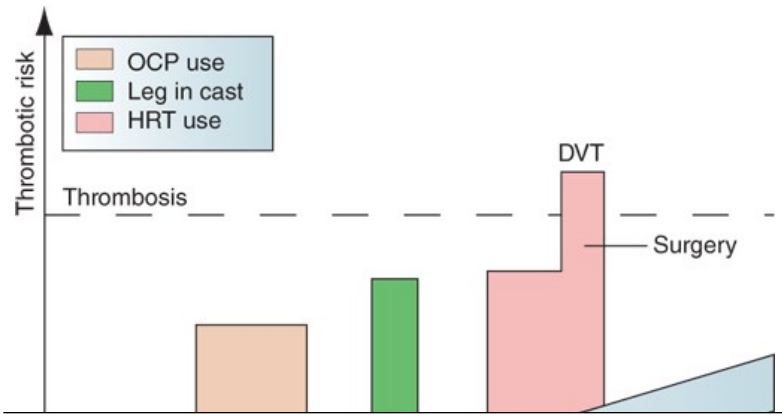
VENOUS	VENOUS AND ARTERIAL
Inherited	**Inherited**
Factor V Leiden	Homocystinuria
Prothrombin G20210A	Dysfibrinogenemia
Antithrombin deficiency	**Acquired**
Protein C deficiency	Malignancy
Protein S deficiency	Antiphospholipid antibody syndrome
Acquired	Hormonal therapy
Age	Polycythemia vera
Previous thrombosis	Essential thrombocythemia
Immobilization	Paroxysmal nocturnal hemoglobinuria
Major surgery	Thrombotic thrombocytopenic purpura
Pregnancy and puerperium	Heparin-induced thrombocytopenia
Hospitalization	Disseminated intravascular coagulation
Obesity	**Unknown**[a]
Infection	Elevated factor II, VIII, IX, XI
Smoking	Elevated TAFI levels
	Low levels of TFPI

[a]Unknown whether risk is inherited or acquired.

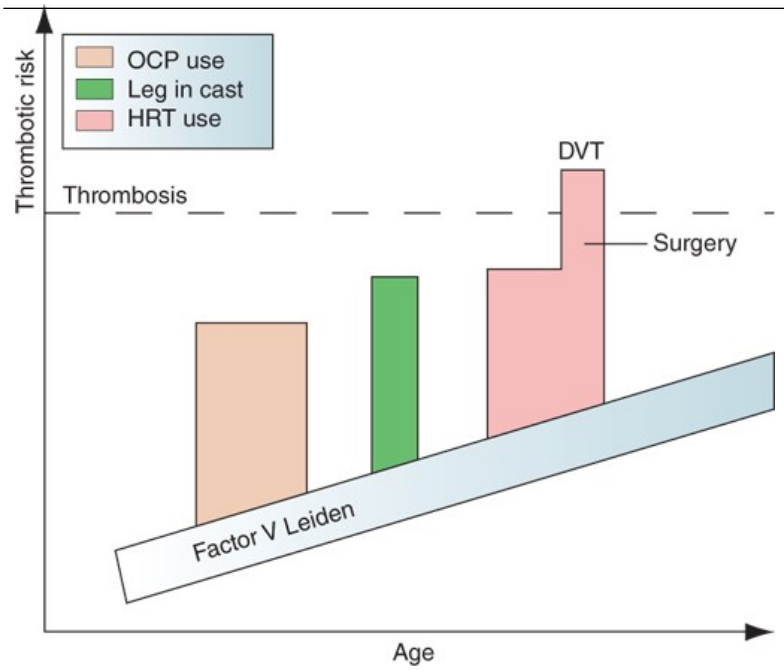
Abbreviations: APC, activated protein C; TAFI, thrombin-activatable fibrinolysis inhibitor; TFPI, tissue factor pathway inhibitor.

FIGURE 65-5

Thrombotic risk over time. Shown schematically is an individual's thrombotic risk over time. An underlying factor V Leiden variant provides a "theoretically" constant increased risk. The thrombotic risk increases with age and, intermittently, with oral contraceptive (OCP) or hormone replacement therapy (HRT) use; other events, like major surgery or illness, will increase the risk further. At some point, the cumulative risk may increase to the threshold for thrombosis and result in deep venous thrombosis (DVT). Note: The magnitude and duration of risk portrayed in the figure are meant for example only and may not precisely reflect the relative risk determined by clinical study. *(Sources: From BA Konkle, A Schafer, in DP Zipes et al [eds]: Braunwald's Heart Disease, 7th ed. Philadelphia, Saunders, 2005; from FR Rosendaal: Venous thrombosis: A multicausal disease. Lancet 353:1167, 1999.)*



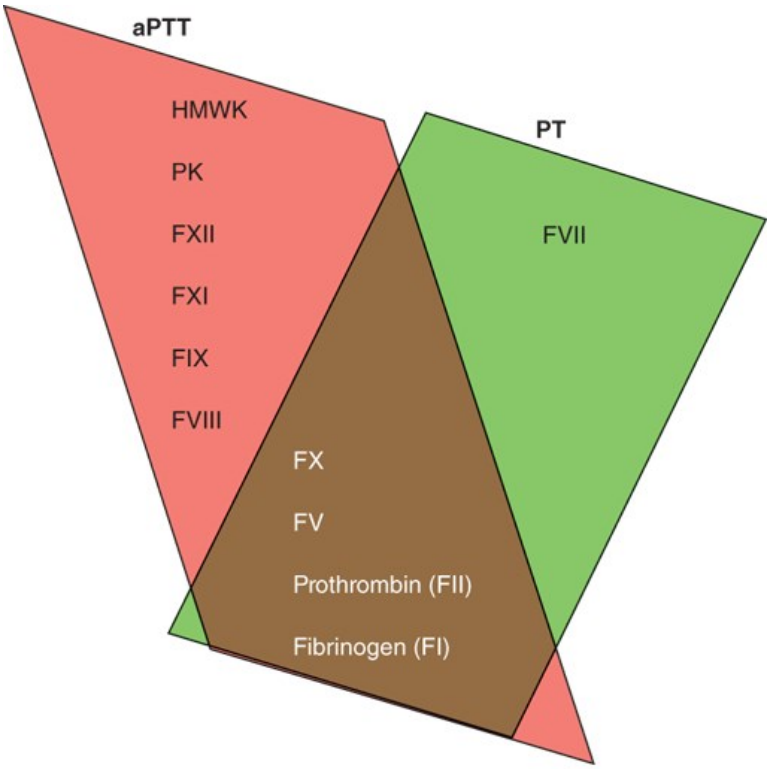
Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

FIGURE 65-6

Coagulation factor activity tested in the activated partial thromboplastin time (aPTT) in red and prothrombin time (PT) in green, or both. F, factor; HMWK, high-molecular-weight kininogen; PK, prekallikrein.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

TABLE 65-4

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Prolonged Activated Partial Thromboplastin Time (aPTT)

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS
Copyright © McGraw Hill. All rights reserved.

TABLE 65-4

Hemostatic Disorders and Coagulation Test Abnormalities

Prolonged Activated Partial Thromboplastin Time (aPTT)
No clinical bleeding—↓ factor XII, high-molecular-weight kininogen, prekallikrein
Variable, but usually mild, bleeding—↓ factor XI, mild ↓ factor VIII and factor IX
Frequent, severe bleeding—severe deficiencies of factors VIII and IX
Heparin and direct thrombin inhibitors
Prolonged Prothrombin Time (PT)
Factor VII deficiency
Vitamin K deficiency—early
Warfarin anticoagulation
Direct Xa inhibitors (rivaroxaban, edoxaban, apixaban—note PT may be normal)
Prolonged aPTT and PT
Factor II, V, X, or fibrinogen deficiency
Vitamin K deficiency—late
Direct thrombin inhibitors
Prolonged Thrombin Time
Heparin or heparin-like inhibitors
Direct thrombin inhibitors (e.g., dabigatran, argatroban, bivalirudin)
Mild or no bleeding—dysfibrinogenemia
Frequent, severe bleeding—afibrinogenemia
Prolonged PT and/or aPTT Not Corrected with Mixing with Normal Plasma
Bleeding—specific factor inhibitor
No symptoms, or clotting and/or pregnancy loss—lupus anticoagulant
Disseminated intravascular coagulation
Heparin or direct thrombin inhibitor
Abnormal Clot Solubility
Rapid Clot Lysis

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Abnormal Clot Solubility
Factor XIII deficiency
Inhibitors or defective cross-linking
Rapid Clot Lysis
Deficiency of α_2-antiplasmin or plasminogen activator inhibitor 1
Treatment with fibrinolytic therapy

FURTHER READING

Chapin JC, Hajjar KA: Fibrinolysis and the control of blood coagulation. Blood Rev 29:17, 2015. [PubMed: 25294122]

Connors JM: Thrombophilia testing and venous thrombosis. N Engl J Med 377:12, 2017.

Connors JM: Testing and monitoring direct oral anticoagulants. Blood 132:2009, 2018. [PubMed: 30201757]

Darzi AJ et al: Prognostic factors for VTE and bleeding in hospitalized medical patients: A systematic review and meta-analysis. Blood 135:1788, 2020. [PubMed: 32092132]

Devreese KMJ et al: Guidance from the Scientific and Standardization Committee for lupus anticoagulant/antiphospholipid antibodies of the International Society on Thrombosis and Haemostasis Update of the guidelines for lupus anticoagulant detection and interpretation. J Thromb Haemost 18:2828, 2020. [PubMed: 33462974]

Elbaz C, Sholzberg M: An illustrated review of bleeding assessment tools and common coagulation tests. Res Pract Thromb Haemost 4:761, 2020. [PubMed: 32685885]

James PD et al: ASH ISTH NHF WFH 2021 guidelines on the diagnosis of von Willebrand disease. Blood Adv 5:280, 2021. [PubMed: 33570651]

Kaufman RM et al: Platelet transfusion: A clinical practice guideline from the AABB. Ann Intern Med 162:205, 2020.

Mackie I et al: Guidelines on the laboratory aspect of assays used in haemostasis and thrombosis. Int J Lab Hem 35:1, 2013.

Moran J, Bauer KA: Managing thromboembolic risk in patients with hereditary and acquired thrombophilias. Blood 135:344, 2020. [PubMed: 31917425]

Wagenman BL et al: The laboratory approach to inherited and acquired coagulation factor deficiencies. Clin Lab Med 29:229, 2009. [PubMed: 19665676]

Yau JW et al: Endothelial cell control of thrombosis. BMC Cardiovasc Disord 15:130, 2015. [PubMed: 26481314]

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 65: Bleeding and Thrombosis, Barbara A. Konkle
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 66: Enlargement of Lymph Nodes and Spleen

Dan L. Longo

INTRODUCTION

This chapter is intended to serve as a guide to the evaluation of patients who present with enlargement of the lymph nodes (*lymphadenopathy*) or the spleen (*splenomegaly*). Lymphadenopathy is a rather common clinical finding in primary care settings, whereas palpable splenomegaly is less so.

LYMPHADENOPATHY

Lymphadenopathy may be an incidental finding in patients being examined for various reasons, or it may be a presenting sign or symptom of the patient's illness. The physician must eventually decide whether the lymphadenopathy is a normal finding or one that requires further study, up to and including biopsy. Soft, flat, submandibular nodes (<1 cm) are often palpable in healthy children and young adults; healthy adults may have palpable inguinal nodes of up to 2 cm, which are considered normal. Further evaluation of these normal nodes is not warranted. In contrast, if the physician believes the node(s) to be abnormal, then pursuit of a more precise diagnosis is needed.

APPROACH TO THE PATIENT

Lymphadenopathy

Lymphadenopathy may be a primary or secondary manifestation of numerous disorders, as shown in **Table 66-1**. Many of these disorders are infrequent causes of lymphadenopathy. In primary care practice, more than two-thirds of patients with lymphadenopathy have nonspecific causes or upper respiratory illnesses (viral or bacterial) and <1% have a malignancy. In one study, 84% of patients referred for evaluation of lymphadenopathy had a "benign" diagnosis. The remaining 16% had a malignancy (lymphoma or metastatic adenocarcinoma). Of the patients with benign lymphadenopathy, 63% had a nonspecific or reactive etiology (no causative agent found), and the remainder had a specific cause demonstrated, most commonly infectious mononucleosis, toxoplasmosis, or tuberculosis. Thus, the vast majority of patients with lymphadenopathy will have a nonspecific etiology requiring few diagnostic tests.

CLINICAL ASSESSMENT

The physician will be aided in the pursuit of an explanation for the lymphadenopathy by a careful medical history, physical examination, selected laboratory tests, and perhaps an excisional lymph node biopsy.

The *medical history* should reveal the setting in which lymphadenopathy is occurring. Symptoms such as sore throat, cough, fever, night sweats, fatigue, weight loss, or pain in the nodes should be sought. The patient's age, sex, occupation, exposure to pets, sexual behavior, and use of drugs such as diphenylhydantoin are other important historic points. For example, children and young adults usually have benign (i.e., nonmalignant) disorders that account for the observed lymphadenopathy such as viral or bacterial upper respiratory infections; infectious mononucleosis; toxoplasmosis; and, in some countries, tuberculosis. In contrast, after age 50, the incidence of malignant disorders increases and that of benign disorders decreases.

The *physical examination* can provide useful clues such as the extent of lymphadenopathy (localized or generalized), size of nodes, texture, presence or absence of nodal tenderness, signs of inflammation over the node, skin lesions, and splenomegaly. A thorough ear, nose, and throat (ENT) examination is indicated in adult patients with cervical adenopathy and a history of tobacco use. Localized or regional adenopathy implies involvement of a single anatomic area. Generalized adenopathy has been defined as involvement of three or more noncontiguous lymph node areas. Many of the causes of lymphadenopathy (Table 66-1) can produce localized *or* generalized adenopathy, so this distinction is of limited utility in the differential diagnosis. Nevertheless, generalized lymphadenopathy is frequently associated with nonmalignant disorders such as infectious

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The site of localized or regional adenopathy may provide a useful clue about the cause. Occipital adenopathy often reflects an infection of the scalp,

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

presence or absence of nodal tenderness, signs of inflammation over the node, skin lesions, and splenomegaly. A thorough ear, nose, and throat in the differential diagnosis. Nevertheless, generalized lymphadenopathy is frequently associated with nonmalignant disorders such as infectious mononucleosis (Epstein-Barr virus [EBV] or cytomegalovirus [CMV]), toxoplasmosis, AIDS, other viral infections, systemic lupus erythematosus (SLE), and mixed connective tissue disease. Acute and chronic lymphocytic leukemias and malignant lymphomas also produce generalized adenopathy in adults.

The site of localized or regional adenopathy may provide a useful clue about the cause. Occipital adenopathy often reflects an infection of the scalp, and preauricular adenopathy accompanies conjunctival infections and cat-scratch disease. The most frequent site of regional adenopathy is the neck, and most of the causes are benign—upper respiratory infections, oral and dental lesions, infectious mononucleosis, or other viral illnesses. The chief malignant causes include metastatic cancer from head and neck, breast, lung, and thyroid primaries. Enlargement of supraclavicular and scalene nodes is always abnormal. Because these nodes drain regions of the lung and retroperitoneal space, they can reflect lymphomas, other cancers, or infectious processes arising in these areas. Virchow's node is an enlarged left supraclavicular node infiltrated with metastatic cancer from a gastrointestinal primary. Metastases to supraclavicular nodes also occur from lung, breast, testis, or ovarian cancers. Tuberculosis, sarcoidosis, and toxoplasmosis are nonneoplastic causes of supraclavicular adenopathy. Axillary adenopathy is usually due to injuries or localized infections of the ipsilateral upper extremity. Malignant causes include melanoma or lymphoma and, in women, breast cancer. Inguinal lymphadenopathy is usually secondary to infections or trauma of the lower extremities and may accompany sexually transmitted diseases such as lymphogranuloma venereum, primary syphilis, genital herpes, or chancroid. These nodes may also be involved by lymphomas and metastatic cancer from primary lesions of the rectum, genitalia, or lower extremities (melanoma).

The size and texture of the lymph node(s) and the presence of pain are useful parameters in evaluating a patient with lymphadenopathy. Nodes <1.0 cm^2 in area (1.0 cm × 1.0 cm or less) are almost always secondary to benign, nonspecific reactive causes. In one retrospective analysis of younger patients (9–25 years) who had a lymph node biopsy, a maximum diameter of >2 cm served as one discriminant for predicting that the biopsy would reveal malignant or granulomatous disease. Another study showed that a lymph node size of 2.25 cm^2 (1.5 cm × 1.5 cm) was the best size limit for distinguishing malignant or granulomatous lymphadenopathy from other causes of lymphadenopathy. Patients with node(s) ≤1.0 cm^2 should be observed after excluding infectious mononucleosis and/or toxoplasmosis unless there are symptoms and signs of an underlying systemic illness.

The texture of lymph nodes may be described as soft, firm, rubbery, hard, discrete, matted, tender, movable, or fixed. Tenderness is found when the capsule is stretched during rapid enlargement, usually secondary to an inflammatory process. Some malignant diseases such as acute leukemia may produce rapid enlargement and pain in the nodes. Nodes involved by lymphoma tend to be large, discrete, symmetric, rubbery, firm, mobile, and nontender. Nodes containing metastatic cancer are often hard, nontender, and nonmovable because of fixation to surrounding tissues. The coexistence of splenomegaly in the patient with lymphadenopathy implies a systemic illness such as infectious mononucleosis, lymphoma, acute or chronic leukemia, SLE, sarcoidosis, toxoplasmosis, cat-scratch disease, or other less common hematologic disorders. The patient's story should provide helpful clues about the underlying systemic illness.

Nonsuperficial presentations (thoracic or abdominal) of adenopathy are usually detected as the result of a symptom-directed diagnostic workup. Thoracic adenopathy may be detected by routine chest radiography or during the workup for superficial adenopathy. It may also be found because the patient complains of a cough or wheezing from airway compression; hoarseness from recurrent laryngeal nerve involvement; dysphagia from esophageal compression; or swelling of the neck, face, or arms secondary to compression of the superior vena cava or subclavian vein. The differential diagnosis of mediastinal and hilar adenopathy includes primary lung disorders and systemic illnesses that characteristically involve mediastinal or hilar nodes. In the young, mediastinal adenopathy is associated with infectious mononucleosis and sarcoidosis. In endemic regions, histoplasmosis can cause unilateral paratracheal lymph node involvement that mimics lymphoma. Tuberculosis can also cause unilateral adenopathy. In older patients, the differential diagnosis includes primary lung cancer (especially among smokers), lymphomas, metastatic carcinoma (usually lung), tuberculosis, fungal infection, and sarcoidosis.

Enlarged intraabdominal or retroperitoneal nodes are usually malignant. Although tuberculosis may present as mesenteric lymphadenitis, these masses usually contain lymphomas or, in young men, germ cell tumors.

LABORATORY INVESTIGATION

The laboratory investigation of patients with lymphadenopathy must be tailored to elucidate the etiology suspected from the patient's history and physical findings. One study from a family practice clinic evaluated 249 younger patients with "enlarged lymph nodes, not infected" or "lymphadenitis." No laboratory studies were obtained in 51%. When studies were performed, the most common were a complete blood count (CBC) (33%), throat culture (16%), chest x-ray (12%), or monospot test (10%). Only eight patients (3%) had a node biopsy, and half of those were normal or reactive. The CBC can provide useful data for the diagnosis of acute or chronic leukemias, EBV or CMV mononucleosis, lymphoma with a leukemic

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The chest x-ray is usually negative, but the presence of a pulmonary infiltrate or mediastinal lymphadenopathy would suggest tuberculosis,

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

(33%), throat culture (16%), chest x-ray (12%), or monospot test (10%). Only eight patients (3%) had a node biopsy, and half of those were normal or reactive. The CBC can provide useful data for the diagnosis of acute or chronic leukemias, EBV or CMV mononucleosis, lymphoma with a leukemic component, pyogenic infections, or immune cytopenias in illnesses such as SLE. Serologic studies may demonstrate antibodies specific to components of EBV, CMV, HIV, and other viruses; *Toxoplasma gondii*; *Brucella*; etc. If SLE is suspected, antinuclear and anti-DNA antibody studies are warranted.

The chest x-ray is usually negative, but the presence of a pulmonary infiltrate or mediastinal lymphadenopathy would suggest tuberculosis, histoplasmosis, sarcoidosis, lymphoma, primary lung cancer, or metastatic cancer and demands further investigation.

A variety of imaging techniques (CT, MRI, ultrasound, color Doppler ultrasonography) have been employed to differentiate benign from malignant lymph nodes, especially in patients with head and neck cancer. CT and MRI are comparably accurate (65–90%) in the diagnosis of metastases to cervical lymph nodes. Ultrasonography has been used to determine the long (L) axis, short (S) axis, and a ratio of long to short axis in cervical nodes. An L/S ratio of <2.0 has a sensitivity and a specificity of 95% for distinguishing benign and malignant nodes in patients with head and neck cancer. This ratio has greater specificity and sensitivity than palpation or measurement of either the long or the short axis alone.

The indications for lymph node biopsy are imprecise, yet it is a valuable diagnostic tool. The decision to biopsy may be made early in a patient's evaluation or delayed for up to 2 weeks. Prompt biopsy should occur if the patient's history and physical findings suggest a malignancy; examples include a solitary, hard, nontender cervical node in an older patient who is a chronic user of tobacco; supraclavicular adenopathy; and solitary or generalized adenopathy that is firm, movable, and suggestive of lymphoma. If a primary head and neck cancer is suspected as the basis of a solitary, hard cervical node, then a careful ENT examination should be performed. Any mucosal lesion that is suspicious for a primary neoplastic process should be biopsied first. If no mucosal lesion is detected, an excisional biopsy of the largest node should be performed. Fine-needle aspiration should not be performed as the first diagnostic procedure. Most diagnoses require more tissue than such aspiration can provide, and it often delays a definitive diagnosis. Fine-needle aspiration should be reserved for thyroid nodules and for confirmation of relapse in patients whose primary diagnosis is known. If the primary physician is uncertain about whether to proceed to biopsy, consultation with a hematologist or medical oncologist should be helpful. In primary care practices, <5% of lymphadenopathy patients will require a biopsy. That percentage will be considerably larger in referral practices, i.e., hematology, oncology, or ENT.

Two groups have reported algorithms that they claim will identify more precisely those lymphadenopathy patients who should have a biopsy. Both reports were retrospective analyses in referral practices. The first study involved patients 9–25 years of age who had a node biopsy performed. Three variables were identified that predicted those young patients with peripheral lymphadenopathy who should undergo biopsy; lymph node size >2 cm in diameter and abnormal chest x-ray had positive predictive values, whereas recent ENT symptoms had negative predictive values. The second study evaluated 220 lymphadenopathy patients in a hematology unit and identified five variables (lymph node size, location [supraclavicular or nonsupraclavicular], age [>40 years or <40 years], texture [nonhard or hard], and tenderness) that were used in a mathematical model to identify those patients requiring a biopsy. Positive predictive value was found for age >40 years, supraclavicular location, node size >2.25 cm^2, hard texture, and lack of pain or tenderness. Negative predictive value was evident for age <40 years, node size <1.0 cm^2, nonhard texture, and tender or painful nodes. Ninety-one percent of those who required biopsy were correctly classified by this model. Because both of these studies were retrospective analyses and one was limited to young patients, it is not known how useful these models would be if applied prospectively in a primary care setting.

Most lymphadenopathy patients do not require a biopsy, and at least half require no laboratory studies. If the patient's history and physical findings point to a benign cause for lymphadenopathy, careful follow-up at a 2- to 4-week interval can be employed. The patient should be instructed to return for reevaluation if there is an increase in the size of the nodes. Antibiotics are not indicated for lymphadenopathy unless strong evidence of a bacterial infection is present. Glucocorticoids should not be used to treat lymphadenopathy because their lympholytic effect obscures some diagnoses (lymphoma, leukemia, Castleman's disease) and they contribute to delayed healing or activation of underlying infections. An exception to this statement is the life-threatening pharyngeal obstruction by enlarged lymphoid tissue in Waldeyer's ring that is occasionally seen in infectious mononucleosis.

TABLE 66-1

Diseases Associated with Lymphadenopathy

1. Infectious diseases
   - Viral—infectious mononucleosis (EBV, CMV), infectious hepatitis, herpesvirus 6, varicella, measles, rubella, rubeola
   - Bacterial—streptococcal pharyngitis, staphylococcal infection, cat-scratch disease, mycobacterial infection, primary and secondary syphilis, diphtheria, leprosy, bartonella
   - Fungal—histoplasmosis, coccidioidomycosis, paracoccidioidomycosis

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Access Provided by:
SILVERCHAIR

1. Infectious diseases
   a. Viral—infectious mononucleosis syndromes (EBV, CMV), infectious hepatitis, herpes simplex, herpesvirus-6, varicella-zoster virus, rubella, measles, adenovirus, HIV, epidemic keratoconjunctivitis, vaccinia, herpesvirus-8
   b. Bacterial—streptococci, staphylococci, cat-scratch disease, brucellosis, tularemia, plague, chancroid, melioidosis, glanders, tuberculosis, atypical mycobacterial infection, primary and secondary syphilis, diphtheria, leprosy, bartonella
   c. Fungal—histoplasmosis, coccidioidomycosis, paracoccidioidomycosis
   d. Chlamydial—lymphogranuloma venereum, trachoma
   e. Parasitic—toxoplasmosis, leishmaniasis, trypanosomiasis, filariasis
   f. Rickettsial—scrub typhus, rickettsialpox, Q fever
2. Immunologic diseases
   a. Rheumatoid arthritis
   b. Juvenile rheumatoid arthritis
   c. Mixed connective tissue disease
   d. Systemic lupus erythematosus
   e. Dermatomyositis
   f. Sjögren's syndrome
   g. Serum sickness
   h. Drug hypersensitivity—diphenylhydantoin, hydralazine, allopurinol, primidone, gold, carbamazepine, etc.
   i. Angioimmunoblastic lymphadenopathy
   j. Primary biliary cirrhosis
   k. Graft-vs-host disease
   l. Silicone-associated
   m. Autoimmune lymphoproliferative syndrome
   n. IgG4-related disease
   o. Immune reconstitution inflammatory syndrome (IRIS)
3. Malignant diseases
   a. Hematologic—Hodgkin's disease, non-Hodgkin's lymphomas, acute or chronic lymphocytic leukemia, hairy cell leukemia, malignant histiocytosis, amyloidosis
   b. Metastatic—from numerous primary sites
4. Lipid storage diseases—Gaucher's, Niemann-Pick, Fabry, Tangier
5. Endocrine diseases—hyperthyroidism
6. Other disorders
   a. Castleman's disease (giant lymph node hyperplasia)
   b. Sarcoidosis
   c. Dermatopathic lymphadenitis
   d. Lymphomatoid granulomatosis
   e. Histiocytic necrotizing lymphadenitis (Kikuchi's disease)
   f. Sinus histiocytosis with massive lymphadenopathy (Rosai-Dorfman disease)
   g. Mucocutaneous lymph node syndrome (Kawasaki's disease)
   h. Histiocytosis X
   i. Familial Mediterranean fever
   j. Severe hypertriglyceridemia
   k. Vascular transformation of sinuses
   l. Inflammatory pseudotumor of lymph node
   m. Congestive heart failure

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

STRUCTURE AND FUNCTION OF THE SPLEEN

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus.

SPLENOMEGALY

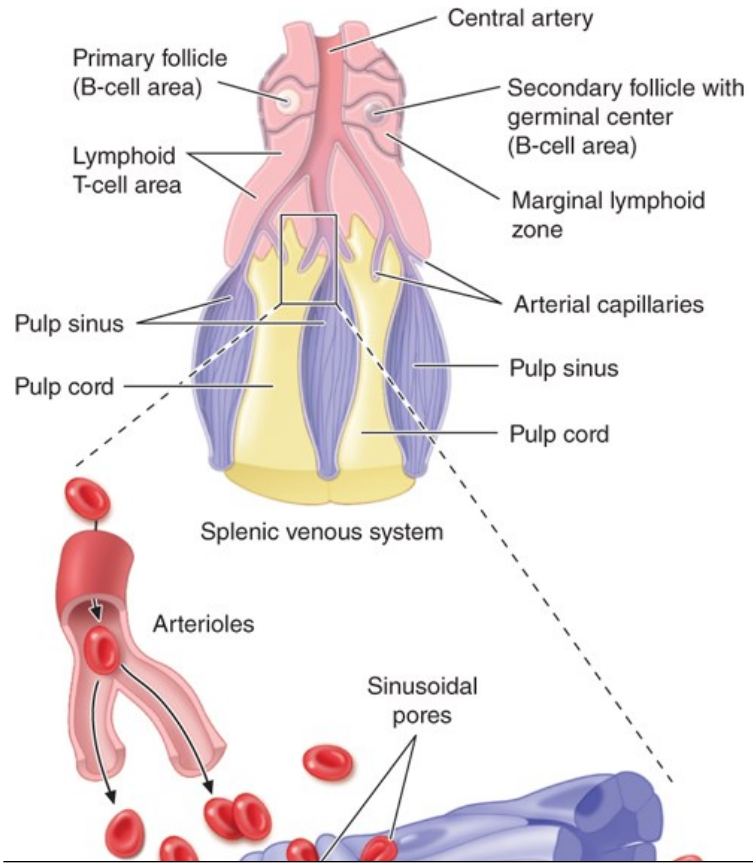
STRUCTURE AND FUNCTION OF THE SPLEEN

The spleen is a reticuloendothelial organ that has its embryologic origin in the dorsal mesogastrium at about 5 weeks' gestation. It arises in a series of hillocks, migrates to its normal adult location in the left upper quadrant (LUQ), and is attached to the stomach via the gastrolienal ligament and to the kidney via the lienorenal ligament. When the hillocks fail to unify into a single tissue mass, accessory spleens may develop in around 20% of persons. The function of the spleen has been elusive. Galen believed it was the source of "black bile" or melancholia, and the word *hypochondria* (literally, beneath the ribs) and the idiom "to vent one's spleen" attest to the beliefs that the spleen had an important influence on the psyche and emotions. In humans, its normal physiologic roles seem to be the following:

1. Maintenance of quality control over erythrocytes in the red pulp by removal of senescent and defective red blood cells. The spleen accomplishes this function through a unique organization of its parenchyma and vasculature **(Fig. 66-1)**.
2. Synthesis of antibodies in the white pulp.
3. The removal of antibody-coated bacteria and antibody-coated blood cells from the circulation.

FIGURE 66-1

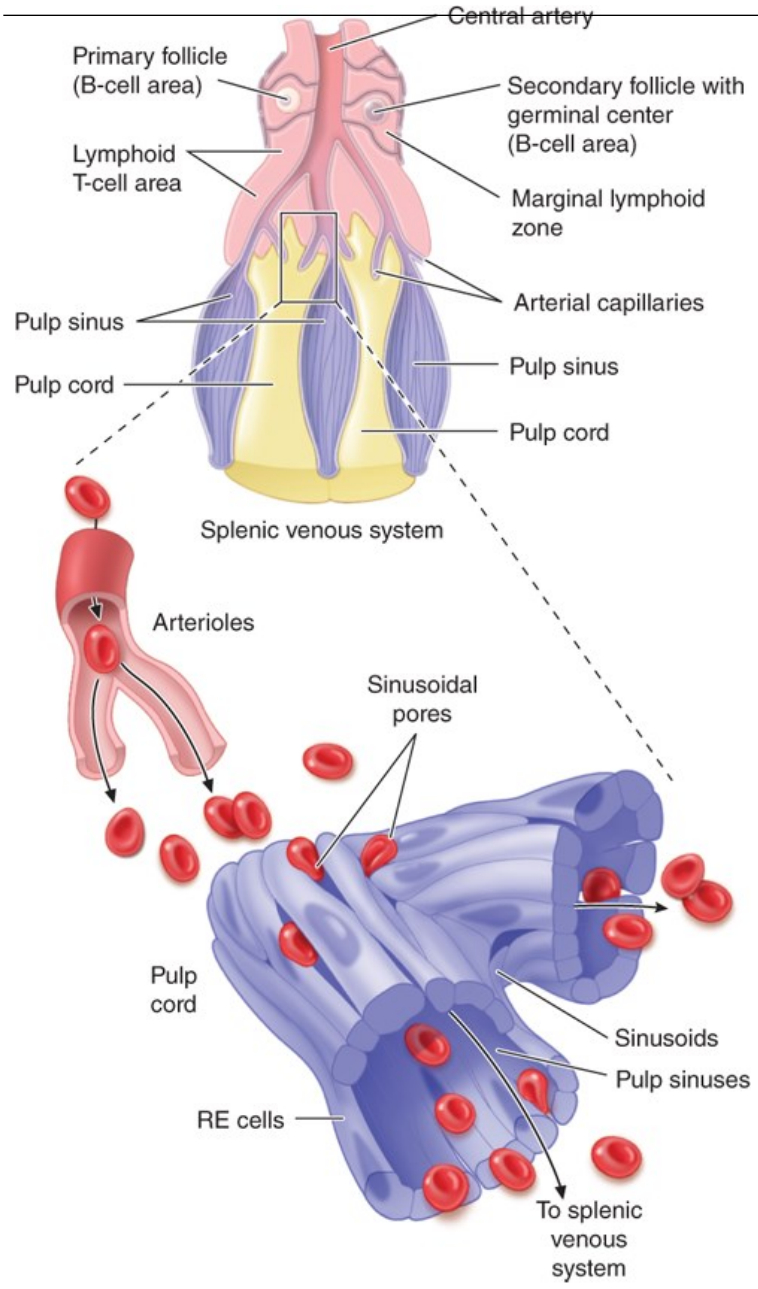
Schematic spleen structure. The spleen comprises many units of red and white pulp centered around small branches of the splenic artery, called *central arteries*. White pulp is lymphoid in nature and contains B-cell follicles, a marginal zone around the follicles, and T-cell–rich areas sheathing arterioles. The red pulp areas include pulp sinuses and pulp cords. The cords are dead ends. In order to regain access to the circulation, red blood cells must traverse tiny openings in the sinusoidal lining. Stiff, damaged, or old red cells cannot enter the sinuses. RE, reticuloendothelial. *(Bottom portion of figure reproduced with permission from RS Hillman, KA Ault, in Hematology in Clinical Practice, 4th ed. New York, McGraw-Hill, 2005.)*



Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility



ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

An increase in these normal functions may result in splenomegaly.

The spleen is composed of *red pulp* and *white pulp*, which are Malpighi's terms for the red blood–filled sinuses and reticuloendothelial cell–lined cords and the white lymphoid follicles arrayed within the red pulp matrix. The spleen is in the portal circulation. The reason for this is unknown but may relate to the fact that lower blood pressure allows less rapid flow and minimizes damage to normal erythrocytes. Blood flows into the spleen at a rate of about 150 mL/min through the splenic artery, which ultimately ramifies into central arterioles. Some blood goes from the arterioles to capillaries and then to splenic veins and out of the spleen, but the majority of blood from central arterioles flows into the macrophage-lined sinuses and cords. The blood entering the sinuses reenters the circulation through the splenic venules, but the blood entering the cords is subjected to an inspection of sorts. To return to the circulation, the blood cells in the cords must squeeze through slits in the cord lining to enter the sinuses that lead to the venules. Old and damaged erythrocytes are less deformable and are retained in the cords, where they are destroyed and their components recycled. Red cell–inclusion bodies such as parasites **(Chaps. 224, 225, and A2)**, nuclear residua (Howell-Jolly bodies, **see Fig. 63-6)**, or denatured hemoglobin (Heinz bodies) are pinched off in the process of passing through the slits, a process called *pitting*. The culling of dead and damaged cells

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

The spleen is also capable of assisting the host in adapting to its hostile environment. It has at least three adaptive functions: (1) clearance of bacteria and particulates from the blood, (2) the generation of immune responses to certain pathogens, and (3) the generation of cellular components of the

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

hemoglobin (Heinz bodies) are pinched off in the process of passing through the slits, a process called *pitting*. The culling of dead and damaged cells and the pitting of cells with inclusions appear to occur without significant delay because the blood transit time through the spleen is only slightly slower than in other organs.

The spleen is also capable of assisting the host in adapting to its hostile environment. It has at least three adaptive functions: (1) clearance of bacteria and particulates from the blood, (2) the generation of immune responses to certain pathogens, and (3) the generation of cellular components of the blood under circumstances in which the marrow is unable to meet the needs (i.e., extramedullary hematopoiesis). The latter adaptation is a recapitulation of the blood-forming function the spleen plays during gestation. In some animals, the spleen also serves a role in the vascular adaptation to stress because it stores red blood cells (often hemoconcentrated to higher hematocrits than normal) under normal circumstances and contracts under the influence of β-adrenergic stimulation to provide the animal with an autotransfusion and improved oxygen-carrying capacity. However, the normal human spleen does not sequester or store red blood cells and does not contract in response to sympathetic stimuli. The normal human spleen contains approximately one-third of the total body platelets and a significant number of marginated neutrophils. These sequestered cells are available when needed to respond to bleeding or infection.

APPROACH TO THE PATIENT

Splenomegaly

CLINICAL ASSESSMENT

The most common *symptoms* produced by diseases involving the spleen are pain and a heavy sensation in the LUQ. Massive splenomegaly may cause early satiety. Pain may result from acute swelling of the spleen with stretching of the capsule, infarction, or inflammation of the capsule. For many years, it was believed that splenic infarction was clinically silent, which, at times, is true. However, Soma Weiss, in his classic 1942 report of the self-observations by a Harvard medical student on the clinical course of subacute bacterial endocarditis, documented that severe LUQ and pleuritic chest pain may accompany thromboembolic occlusion of splenic blood flow. Vascular occlusion, with infarction and pain, is commonly seen in children with sickle cell crises. Rupture of the spleen, from either trauma or infiltrative disease that breaks the capsule, may result in intraperitoneal bleeding, shock, and death. The rupture itself may be painless.

A palpable spleen is the major *physical sign* produced by diseases affecting the spleen and suggests enlargement of the organ. The normal spleen weighs <250 g, decreases in size with age, normally lies entirely within the rib cage, has a maximum cephalocaudad diameter of 13 cm by ultrasonography or maximum length of 12 cm and/or width of 7 cm by radionuclide scan, and is usually not palpable. However, a palpable spleen was found in 3% of 2200 asymptomatic, male, freshman college students. Follow-up at 3 years revealed that 30% of those students still had a palpable spleen without any increase in disease prevalence. Ten-year follow-up found no evidence for lymphoid malignancies. Furthermore, in some tropical countries (e.g., New Guinea), the incidence of splenomegaly may reach 60%. Thus, the presence of a palpable spleen does not always equate with presence of disease. Even when disease is present, splenomegaly may not reflect the primary disease but rather a reaction to it. For example, in patients with Hodgkin's disease, only two-thirds of the palpable spleens show involvement by the cancer.

Physical examination of the spleen uses primarily the techniques of palpation and percussion. Inspection may reveal fullness in the LUQ that descends on inspiration, a finding associated with a massively enlarged spleen. Auscultation may reveal a venous hum or friction rub.

Palpation can be accomplished by bimanual palpation, ballotment, and palpation from above (Middleton maneuver). For bimanual palpation, which is at least as reliable as the other techniques, the patient is supine with flexed knees. The examiner's left hand is placed on the lower rib cage and pulls the skin toward the costal margin, allowing the fingertips of the right hand to feel the tip of the spleen as it descends while the patient inspires slowly, smoothly, and deeply. Palpation is begun with the right hand in the left lower quadrant with gradual movement toward the left costal margin, thereby identifying the lower edge of a massively enlarged spleen. When the spleen tip is felt, the finding is recorded as centimeters below the left costal margin at some arbitrary point, i.e., 10–15 cm, from the midpoint of the umbilicus or the xiphisternal junction. This allows other examiners to compare findings or the initial examiner to determine changes in size over time. Bimanual palpation in the right lateral decubitus position adds nothing to the supine examination.

Percussion for splenic dullness is accomplished with any of three techniques described by Nixon, Castell, or Barkun:

1. *Nixon's method*: The patient is placed on the right side so that the spleen lies above the colon and stomach. Percussion begins at the lower level of pulmonary resonance in the posterior axillary line and proceeds diagonally along a perpendicular line toward the lower midanterior costal margin. The upper border of dullness is normally 6–8 cm above the costal margin. Dullness >8 cm in an adult is presumed to indicate splenic note if the spleen is normal in size. This is true during expiration or full inspiration. A dull percussion note on full inspiration suggests splenomegaly.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

of pulmonary resonance in the posterior axillary line and proceeds diagonally along a perpendicular line toward the lower midanterior costal margin. The upper border of dullness is normally 6–8 cm above the costal margin. Dullness >8 cm in an adult is presumed to indicate splenic enlargement.

2. *Castell's method*: With the patient supine, percussion in the lowest intercostal space in the anterior axillary line (8th or 9th) produces a resonant note if the spleen is normal in size. This is true during expiration or full inspiration. A dull percussion note on full inspiration suggests splenomegaly.

3. *Percussion of Traube's semilunar space*: The borders of Traube's space are the sixth rib superiorly, the left midaxillary line laterally, and the left costal margin inferiorly. The patient is supine with the left arm slightly abducted. During normal breathing, this space is percussed from medial to lateral margins, yielding a normal resonant sound. A dull percussion note suggests splenomegaly.

Studies comparing methods of percussion and palpation with a standard of ultrasonography or scintigraphy have revealed sensitivity of 56–71% for palpation and 59–82% for percussion. Reproducibility among examiners is better for palpation than percussion. Both techniques are less reliable in obese patients or patients who have just eaten. Thus, the physical examination techniques of palpation and percussion are imprecise at best. It has been suggested that the examiner perform percussion first and, if positive, proceed to palpation; if the spleen is palpable, then one can be reasonably confident that splenomegaly exists. However, not all LUQ masses are enlarged spleens; gastric or colon tumors and pancreatic or renal cysts or tumors can mimic splenomegaly.

The presence of an enlarged spleen can be more precisely determined, if necessary, by liver-spleen radionuclide scan, CT, MRI, or ultrasonography. The latter technique is the current procedure of choice for routine assessment of spleen size (normal = a maximum cephalocaudad diameter of 13 cm) because it has high sensitivity and specificity and is safe, noninvasive, quick, mobile, and less costly. Equipment advances allow ultrasonography to be performed at the bedside with excellent sensitivity and specificity. Nuclear medicine scans are accurate, sensitive, and reliable but are costly, require greater time to generate data, and use immobile equipment. They have the advantage of demonstrating accessory splenic tissue. CT and MRI provide accurate determination of spleen size, but the equipment is immobile and the procedures are expensive. MRI appears to offer no advantage over CT. Changes in spleen structure such as mass lesions, infarcts, inhomogeneous infiltrates, and cysts are more readily assessed by CT, MRI, or ultrasonography. None of these techniques is very reliable in the detection of patchy infiltration (e.g., Hodgkin's disease).

DIFFERENTIAL DIAGNOSIS

Many of the diseases associated with splenomegaly are listed in **Table 66-2**. They are grouped according to the presumed basic mechanisms responsible for organ enlargement:

1. Hyperplasia or hypertrophy related to a particular splenic function such as reticuloendothelial hyperplasia (work hypertrophy) in diseases such as hereditary spherocytosis or thalassemia syndromes that require removal of large numbers of defective red blood cells; immune hyperplasia in response to systemic infection (infectious mononucleosis, subacute bacterial endocarditis) or to immunologic diseases (immune thrombocytopenia, SLE, Felty's syndrome).

2. Passive congestion due to decreased blood flow from the spleen in conditions that produce portal hypertension (cirrhosis, Budd-Chiari syndrome, congestive heart failure).

3. Infiltrative diseases of the spleen (lymphomas, metastatic cancer, amyloidosis, Gaucher's disease, myeloproliferative disorders with extramedullary hematopoiesis).

The differential diagnostic possibilities are much fewer when the spleen is "massively enlarged" or palpable >8 cm below the left costal margin or its drained weight is ≥1000 g **(Table 66-3)**. The vast majority of such patients will have non-Hodgkin's lymphoma, chronic lymphocytic leukemia, hairy cell leukemia, chronic myeloid leukemia, myelofibrosis with myeloid metaplasia, or polycythemia vera.

LABORATORY ASSESSMENT

The major laboratory abnormalities accompanying splenomegaly are determined by the underlying systemic illness. Erythrocyte counts may be normal, decreased (thalassemia major syndromes, SLE, cirrhosis with portal hypertension), or increased (polycythemia vera). Granulocyte counts may be normal, decreased (Felty's syndrome, congestive splenomegaly, leukemias), or increased (infections or inflammatory disease, myeloproliferative disorders). Similarly, the platelet count may be normal, decreased when there is enhanced sequestration or destruction of platelets in an enlarged spleen (congestive splenomegaly, Gaucher's disease, immune thrombocytopenia), or increased in the myeloproliferative

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

splenomegaly, cytopenia(s), normal or hyperplastic bone marrow, and a response to splenectomy. The latter characteristic is less precise because reversal of cytopenia, particularly granulocytopenia, is sometimes not sustained after splenectomy. The cytopenias result from increased

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

myeloproliferative disorders). Similarly, the platelet count may be normal, decreased when there is enhanced sequestration or destruction of platelets in an enlarged spleen (congestive splenomegaly, Gaucher's disease, immune thrombocytopenia), or increased in the myeloproliferative disorders such as polycythemia vera.

The CBC may reveal cytopenia of one or more blood cell types, which should suggest *hypersplenism*. This condition is characterized by splenomegaly, cytopenia(s), normal or hyperplastic bone marrow, and a response to splenectomy. The latter characteristic is less precise because reversal of cytopenia, particularly granulocytopenia, is sometimes not sustained after splenectomy. The cytopenias result from increased destruction of the cellular elements secondary to reduced flow of blood through enlarged and congested cords (congestive splenomegaly) or to immune-mediated mechanisms. In hypersplenism, various cell types usually have normal morphology on the peripheral blood smear, although the red cells may be spherocytic due to loss of surface area during their longer transit through the enlarged spleen. The increased marrow production of red cells should be reflected as an increased reticulocyte production index, although the value may be less than expected due to increased sequestration of reticulocytes in the spleen.

The need for additional laboratory studies is dictated by the differential diagnosis of the underlying illness of which splenomegaly is a manifestation.

TABLE 66-2

Diseases Associated with Splenomegaly Grouped by Pathogenic Mechanism

Enlargement Due to Increased Demand for Splenic Function	
Reticuloendothelial system hyperplasia (for removal of defective erythrocytes)	Leishmaniasis
Spherocytosis	Trypanosomiasis
Early sickle cell anemia	Ehrlichiosis
Ovalocytosis	Disordered immunoregulation
Thalassemia major	Hemophagocytic lymphohistiocytosis (HLH)
Hemoglobinopathies	Rheumatoid arthritis (Felty's syndrome)
Paroxysmal nocturnal hemoglobinuria	Systemic lupus erythematosus
Pernicious anemia	Collagen vascular diseases
Immune hyperplasia	Serum sickness
Response to infection (viral, bacterial, fungal, parasitic)	Immune hemolytic anemias
Infectious mononucleosis	Immune thrombocytopenias
AIDS	Immune neutropenias
Viral hepatitis	Drug reactions
Cytomegalovirus	Angioimmunoblastic lymphadenopathy
Subacute bacterial endocarditis	Sarcoidosis
Splenic abscess	Extramedullary hematopoiesis

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Viral hepatitis	Drug reactions
Subacute bacterial endocarditis	Sarcoidosis
Bacterial septicemia	Thyrotoxicosis (benign lymphoid hypertrophy)
Congenital syphilis	Interleukin 2 therapy
Splenic abscess	Extramedullary hematopoiesis
Tuberculosis	Myelofibrosis
Histoplasmosis	Marrow damage by toxins, radiation, strontium
Malaria	Marrow infiltration by tumors, leukemias, Gaucher's disease
Enlargement Due to Abnormal Splenic or Portal Blood Flow	
Cirrhosis	Splenic artery aneurysm
Hepatic vein obstruction	Hepatic schistosomiasis
Portal vein obstruction, intrahepatic or extrahepatic	Congestive heart failure
Cavernous transformation of the portal vein	Hepatic echinococcosis
Splenic vein obstruction	Portal hypertension (any cause including the above): "Banti's disease"
Infiltration of the Spleen	
Intracellular or extracellular depositions	Hodgkin's disease
Amyloidosis	Myeloproliferative syndromes (e.g., polycythemia vera, essential thrombocytosis)
Gaucher's disease	Angiosarcomas
Niemann-Pick disease	Metastatic tumors (melanoma is most common)
Tangier disease	Eosinophilic granuloma
Hurler's syndrome and other mucopolysaccharidoses	Histiocytosis X
Hyperlipidemias	Hamartomas
Benign and malignant cellular infiltrations	Hemangiomas, fibromas, lymphangiomas
Leukemias (acute, chronic, lymphoid, myeloid, monocytic)	Splenic cysts
Lymphomas	
Unknown Etiology	
Idiopathic splenomegaly	Iron-deficiency anemia
Berylliosis	

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Diseases Associated with Massive Splenomegaly[a]

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

Berylliosis	

TABLE 66-3

Diseases Associated with Massive Splenomegaly[a]

Chronic myeloid leukemia	Gaucher's disease
Lymphomas	Chronic lymphocytic leukemia
Hairy cell leukemia	Sarcoidosis
Myelofibrosis with myeloid metaplasia	Autoimmune hemolytic anemia
Polycythemia vera	Diffuse splenic hemangiomatosis

[a]The spleen extends >8 cm below left costal margin and/or weighs >1000 g.

SPLENECTOMY

Splenectomy is infrequently performed for diagnostic purposes, especially in the absence of clinical illness or other diagnostic tests that suggest underlying disease. More often, splenectomy is performed for symptom control in patients with massive splenomegaly, for disease control in patients with traumatic splenic rupture, or for correction of cytopenias in patients with hypersplenism or immune-mediated destruction of one or more cellular blood elements. Splenectomy is necessary for staging of patients with Hodgkin's disease only in those with clinical stage I or II disease in whom radiation therapy alone is contemplated as the treatment. Noninvasive staging of the spleen in Hodgkin's disease is not a sufficiently reliable basis for treatment decisions because one-third of normal-sized spleens will be involved with Hodgkin's disease and one-third of enlarged spleens will be tumor-free. The widespread use of systemic therapy to test all stages of Hodgkin's disease has made staging laparotomy with splenectomy unnecessary. Although splenectomy in chronic myeloid leukemia (CML) does not affect the natural history of disease, removal of the massive spleen usually makes patients significantly more comfortable and simplifies their management by significantly reducing transfusion requirements. The improvements in therapy of CML have reduced the need for splenectomy for symptom control. Splenectomy is an effective secondary or tertiary treatment for two chronic B-cell leukemias, hairy cell leukemia and prolymphocytic leukemia, and for the very rare splenic mantle cell or marginal zone lymphoma. Splenectomy in these diseases may be associated with significant tumor regression in bone marrow and other sites of disease. Similar regressions of systemic disease have been noted after splenic irradiation in some types of lymphoid tumors, especially chronic lymphocytic leukemia and prolymphocytic leukemia. This has been termed the *abscopal effect.* Such systemic tumor responses to local therapy directed at the spleen suggest that some hormone or growth factor produced by the spleen may affect tumor cell proliferation, but this conjecture is not yet substantiated. A common therapeutic indication for splenectomy is traumatic or iatrogenic splenic rupture. In a fraction of patients with splenic rupture, peritoneal seeding of splenic fragments can lead to *splenosis*—the presence of multiple rests of spleen tissue not connected to the portal circulation. This ectopic spleen tissue may cause pain or gastrointestinal obstruction, as in endometriosis. A large number of hematologic, immunologic, and congestive causes of splenomegaly can lead to destruction of one or more cellular blood elements. In the majority of such cases, splenectomy can correct the cytopenias, particularly anemia and thrombocytopenia. In a large series of patients seen in two tertiary care centers, the indication for splenectomy was diagnostic in 10% of patients, therapeutic in 44%, staging for Hodgkin's disease in 20%, and incidental to another procedure in 26%. Perhaps the only contraindication to splenectomy is the presence of marrow failure, in which the enlarged spleen is the only source of hematopoietic tissue.

Often the splenectomy is done by laparoscopy, which is associated with shorter hospital stays and faster recovery than the open procedure; however, concern has emerged that the laparoscopic approach is associated with a higher risk of postoperative portal venous system thrombosis and Budd-Chiari syndrome.

The absence of the spleen has minimal long-term effects on the hematologic profile. In the immediate postsplenectomy period, leukocytosis (up to 25,000/μL) and thrombocytosis (up to 1×10^6/μL) may develop, but within 2–3 weeks, blood cell counts and survival of each cell lineage are usually

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

splenic infiltration by tumor that has interfered with its normal culling and pitting function.

The most serious consequence of splenectomy is increased susceptibility to bacterial infections, particularly those with capsules such as

25,000/μL) and thrombocytosis (up to 1×10^6/μL) may develop, but within 2–3 weeks, blood cell counts and survival of each cell lineage are usually normal. The chronic manifestations of splenectomy are marked variation in size and shape of erythrocytes (anisocytosis, poikilocytosis) and the presence of Howell-Jolly bodies (nuclear remnants), Heinz bodies (denatured hemoglobin), basophilic stippling, and an occasional nucleated erythrocyte in the peripheral blood. When such erythrocyte abnormalities appear in a patient whose spleen has not been removed, one should suspect splenic infiltration by tumor that has interfered with its normal culling and pitting function.

The most serious consequence of splenectomy is increased susceptibility to bacterial infections, particularly those with capsules such as *Streptococcus pneumoniae*, *Haemophilus influenzae*, and some gram-negative enteric organisms. Patients aged <20 years are particularly susceptible to overwhelming sepsis with *S. pneumoniae*, and the overall actuarial risk of sepsis in patients who have had their spleens removed is about 7% in 10 years. The case-fatality rate for pneumococcal sepsis in splenectomized patients is 50–80%. About 25% of patients without spleens will develop a serious infection at some time in their life. The frequency is highest within the first 3 years after splenectomy. About 15% of the infections are polymicrobial, and lung, skin, and blood are the most common sites. No increased risk of viral infection has been noted in patients who have no spleen. The susceptibility to bacterial infections relates to the inability to remove opsonized bacteria from the bloodstream and a defect in making antibodies to T-cell–independent antigens such as the polysaccharide components of bacterial capsules. Pneumococcal vaccine should be administered to all patients 2 weeks before elective splenectomy. The Advisory Committee on Immunization Practices recommends that these patients receive repeat vaccination 5 years after splenectomy. Efficacy has not been proven for this group, and the recommendation discounts the possibility that administration of the vaccine may actually lower the titer of specific pneumococcal antibodies. A more effective pneumococcal conjugate vaccine that involves T cells in the response is now available (PCV13). The vaccine to *Neisseria meningitidis* should also be given to patients in whom elective splenectomy is planned. Although efficacy data for *Haemophilus influenzae* type b vaccine are not available for older children or adults, it may be given to patients who have had a splenectomy.

Splenectomized patients should be educated to consider any unexplained fever as a medical emergency. Prompt medical attention with evaluation and treatment of suspected bacteremia may be lifesaving. Routine chemoprophylaxis with oral penicillin can result in the emergence of drug-resistant strains and is not recommended.

In addition to an increased susceptibility to bacterial infections, splenectomized patients are also more susceptible to the parasitic disease babesiosis. The splenectomized patient should avoid areas where the parasite *Babesia* is endemic (e.g., Cape Cod, MA).

Surgical removal of the spleen is an obvious cause of hyposplenism. Patients with sickle cell disease often suffer from autosplenectomy as a result of splenic destruction by the numerous infarcts associated with sickle cell crises during childhood. Indeed, the presence of a palpable spleen in a patient with sickle cell disease after age 5 suggests a coexisting hemoglobinopathy, e.g., thalassemia or hemoglobin C. In addition, patients who receive splenic irradiation for a neoplastic or autoimmune disease are also functionally hyposplenic. The term *hyposplenism* is preferred to *asplenism* in referring to the physiologic consequences of splenectomy because asplenia is a rare, specific, and fatal congenital abnormality in which there is a failure of the left side of the coelomic cavity (which includes the splenic anlagen) to develop normally. Infants with asplenia have no spleens, but that is the least of their problems. The right side of the developing embryo is duplicated on the left so there is liver where the spleen should be, there are two right lungs, and the heart comprises two right atria and two right ventricles.

ACKNOWLEDGMENT

Patrick H. Henry, MD, friend and mentor now deceased, contributed significantly to the chapter in past editions, and much of his work remains in this chapter.

FURTHER READING

Barkun AN et al: The bedside assessment of splenic enlargement. Am J Med 91:512, 1991. [PubMed: 1951414]

Cessford T et al: Comparing physical examination with sonographic versions of the same examination techniques for splenomegaly. J Ultrasound Med 37:1621, 2018. [PubMed: 29219201]

Facchetti F: Tumors of the spleen. Int J Surg Pathol 18:136S, 2010. [PubMed: 20484279]

Girard E et al: Management of splenic and pancreatic trauma. J Visc Surg 153(suppl 4):45, 2016. [PubMed: 27402320]

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Kim DK et al: Advisory committee on immunization practices reocommended immunization schedule for adults aged 19 years or older—United States, 2017. MMWR 66:136, 2017. [PubMed: 28182599]

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

Girard E et al: Management of splenic and pancreatic trauma. J Visc Surg 153(suppl 4):45, 2016. [PubMed: 27402320]

Graves SA et al: Does this patient have splenomegaly? JAMA 270:2218, 1993. [PubMed: 8411607]

Kim DK et al: Advisory committee on immunization practices reocommended immunization schedule for adults aged 19 years or older—United States, 2017. MMWR 66:136, 2017. [PubMed: 28182599]

Kraus MD et al: The spleen as a diagnostic specimen: A review of ten years' experience at two tertiary care institutions. Cancer 91:2001, 2001. [PubMed: 11391578]

McIntyre OR, Ebaugh FG Jr: Palpable spleens: Ten-year follow-up. Ann Intern Med 90:130, 1979. [PubMed: 420452]

Pangalis GA et al: Clinical approach to lymphadenopathy. Semin Oncol 20:570, 1993. [PubMed: 8296196]

Williamson HA Jr: Lymphadenopathy in a family practice: A descriptive study of 240 cases. J Fam Pract 20:449, 1985. [PubMed: 3989485]

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 66: Enlargement of Lymph Nodes and Spleen, Dan L. Longo
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 67: Principles of Clinical Pharmacology

Dan M. Roden

INTRODUCTION

Drugs are the cornerstone of modern therapeutics. Nevertheless, it is well recognized among health care providers and the lay community that the outcome of drug therapy varies widely among individuals. While this variability has been perceived as an unpredictable, and therefore inevitable, accompaniment of drug therapy, this is not the case.

Drugs interact with specific target molecules to produce their beneficial and adverse effects. The chain of events between administration of a drug and production of these effects in the body can be divided into two components, both of which contribute to variability in drug actions. The first component comprises the processes that determine drug delivery to, and removal from, molecular targets. The resulting description of the relationship between drug concentration and time is termed *pharmacokinetics*. The second component of variability in drug action comprises the processes that determine variability in drug actions independent of variability in drug delivery to effector drug sites. This description of the relationship between drug concentration and effect is termed *pharmacodynamics*. As discussed further below, pharmacodynamic variability can arise as a result of variability in function of the target molecule itself or of variability in the broad biologic context in which the drug-target interaction occurs to achieve drug effects. The principles described below were developed by studying small drug molecules but are equally useful in describing the effects of very large molecules, such as the therapeutic antibodies increasingly applied to autoimmune diseases and cancer.

Two important goals of clinical pharmacology are (1) to provide a description of conditions under which drug actions vary among human subjects; and (2) to determine mechanisms underlying this variability, with the goal of improving therapy with available drugs as well as pointing to mechanisms whose targeting by new drugs may be effective in the treatment of human disease. The drug development process is briefly described at the end of this chapter.

The first steps in the discipline of clinical pharmacology were empirical descriptions of the influence of disease on drug actions and of individuals or families with unusual sensitivities to adverse drug reactions (ADRs). These important descriptive findings are now being replaced by an understanding of the molecular mechanisms underlying variability in drug actions. Importantly, it is often the personal interaction of the patient with the physician or other health care provider that first identifies unusual variability in drug actions; maintained alertness to unusual drug responses continues to be a key component of improving drug safety.

One useful unifying framework is to consider that the effects of disease, drug coadministration, or familial factors in modulating drug action reflect variability in expression or function of specific genes whose products determine pharmacokinetics and pharmacodynamics. This idea forms the basis for pharmacogenomic science; a few examples are cited in this chapter, and further details are addressed in **Chap. 68**.

GLOBAL CONSIDERATIONS

It is true across all cultures and diseases that factors such as compliance, genetic variants affecting pharmacokinetics or pharmacodynamics (which themselves vary by ancestry), and drug interactions contribute to drug responses. Cost issues or cultural factors may determine the likelihood that specific drugs, drug combinations, or over-the-counter (OTC) remedies are prescribed. The broad principles of clinical pharmacology enunciated here can be used to analyze the mechanisms underlying successful or unsuccessful therapy with any drug.

INDICATIONS FOR DRUG THERAPY: RISK VERSUS BENEFIT

It is self-evident that the benefits of drug therapy should outweigh the risks. Benefits fall into broad categories: alleviation of symptoms, prevention of disease progression or complications, and prolonged life. However, establishing the balance between risk and benefit for an individual patient is not always simple. In addition to variability seen even within highly controlled drug trials, patients treated in clinical settings may display responses that

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Adverse Effects

ACCESS Medicine
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

disease progression or complications, and prolonged life. However, establishing the balance between risk and benefit for an individual patient is not always simple. In addition to variability seen even within highly controlled drug trials, patients treated in clinical settings may display responses that were not observed in trials, sometimes due to comorbidities that were trial exclusion criteria. In addition, therapies that provide symptomatic benefits but shorten life may be entertained in patients with serious and highly symptomatic diseases such as heart failure or cancer. These considerations illustrate the continuing, highly personal nature of the relationship between the prescriber and the patient.

Adverse Effects

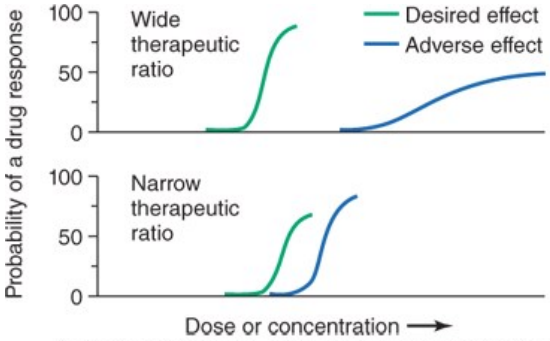
Some adverse effects are so common and so readily associated with drug therapy that they are identified very early during clinical use of a drug. By contrast, serious ADRs may be sufficiently uncommon that they escape detection for many years after a drug begins to be widely used. The issue of how to identify rare but serious ADRs (that can profoundly affect the benefit-risk perception in an individual patient) has not been satisfactorily resolved. Potential approaches range from an increased understanding of the molecular and genetic basis of variability in drug actions to expanded postmarketing surveillance mechanisms. None of these have been completely effective, so practitioners must be continuously vigilant to the possibility that unusual symptoms may be related to specific drugs, or combinations of drugs, that their patients receive.

Therapeutic Index

Beneficial and adverse reactions to drug therapy can be described by a series of dose-response relations **(Fig. 67-1)**. Well-tolerated drugs demonstrate a wide margin, termed the *therapeutic ratio*, *therapeutic index*, or *therapeutic window*, between the doses required to produce a therapeutic effect and those producing toxicity. In cases where there is a similar relationship between plasma drug concentration and effects, monitoring plasma concentrations can be a highly effective aid in managing drug therapy by enabling concentrations to be maintained above the minimum required to produce an effect and below the concentration range likely to produce toxicity. Such monitoring has been widely used to guide therapy with specific agents, such as certain antiarrhythmics, anticonvulsants, and antibiotics. Many of the principles in clinical pharmacology and examples outlined below, which can be applied broadly to therapeutics, have been developed in these arenas.

FIGURE 67-1

The concept of a therapeutic ratio. Each panel illustrates the relationship between increasing dose and cumulative probability of a desired or adverse drug effect. ***Top.*** A drug with a wide therapeutic ratio, that is, a wide separation of the two curves. ***Bottom.*** A drug with a narrow therapeutic ratio; here, the likelihood of adverse effects at therapeutic doses is increased because the curves are not well separated. Further, a steep dose-response curve for adverse effects is especially undesirable, as it implies that even small dosage increments may sharply increase the likelihood of toxicity. When there is a definable relationship between drug concentration (usually measured in plasma) and desirable and adverse effect curves, concentration may be substituted on the abscissa. Note that not all patients necessarily demonstrate a therapeutic response (or adverse effect) at any dose and that some effects (notably some adverse effects) may occur in a dose-independent fashion.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

PRINCIPLES OF PHARMACOKINETICS

The processes of absorption, distribution, metabolism, and excretion—collectively termed *drug disposition*—determine the concentration of drug delivered to target effector molecules.

ABSORPTION AND BIOAVAILABILITY

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

circulation by other routes is termed *bioavailability*. Bioavailability may be <100% for two main reasons: (1) incomplete absorption, or (2) metabolism

ACCESS Medicine

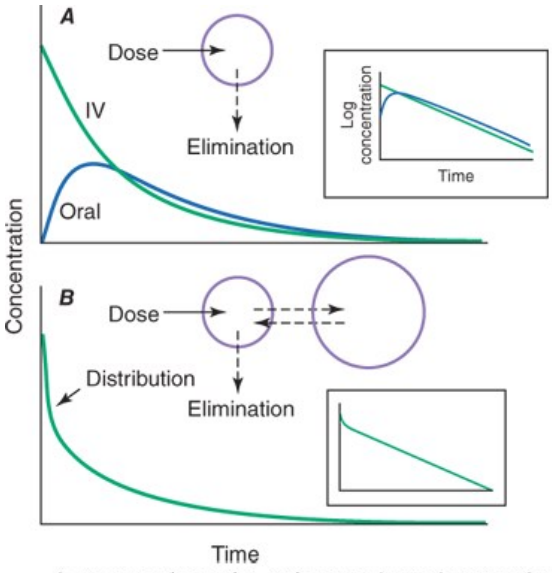
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

ABSORPTION AND BIOAVAILABILITY

When a drug is administered orally, subcutaneously, intramuscularly, rectally, sublingually, or directly into desired sites of action, the amount of drug eventually entering the systemic circulation may be less than with the intravenous route **(Fig. 67-2A)**. The fraction of drug available to the systemic circulation by other routes is termed *bioavailability*. Bioavailability may be <100% for two main reasons: (1) incomplete absorption, or (2) metabolism or elimination prior to entering the systemic circulation.

FIGURE 67-2

Idealized time-plasma concentration curves after a single dose of drug. *A.* The time course of drug concentration after an instantaneous intravenous (IV) bolus or an oral dose in the one-compartment model shown. The area under the time-concentration curve is clearly less with the oral drug than the IV drug, indicating incomplete bioavailability. Note that despite this incomplete bioavailability, concentration after the oral dose can be higher than after the IV dose at some time points. The inset shows that the decline of concentrations over time is linear on a log-linear plot, characteristic of first-order elimination, and that oral and IV drugs have the same elimination (parallel) time course. ***B.*** The decline of central compartment concentration when drug is distributed both to and from a peripheral compartment and eliminated from the central compartment. The rapid initial decline of concentration reflects not drug elimination but distribution.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Compared to the same dose given intravenously, a nonintravenous dose will have a later and lower peak plasma concentration **(Fig. 67-2)**. Drug absorption may be reduced because a drug is incompletely released from its dosage form, undergoes destruction at the site of administration, or has physicochemical properties such as insolubility that prevent complete absorption from its site of administration. Slow absorption rates are deliberately designed into "slow-release" or "sustained-release" drug formulations in order to minimize variation in plasma concentrations during the interval between doses. Therapeutic antibodies administered subcutaneously may take days to reach the systemic circulation.

"First-Pass" Effect

When a drug is administered orally, it must traverse the intestinal epithelium, the portal venous system, and the liver prior to entering the systemic circulation **(Fig. 67-3)**. Once a drug enters the enterocyte, it may undergo metabolism, be transported into the portal vein, or be excreted back into the intestinal lumen. Both excretion into the intestinal lumen and metabolism decrease bioavailability. Once a drug passes this enterocyte barrier, it may also be taken up into the hepatocyte, where bioavailability can be further limited by metabolism or excretion into the bile. This elimination in intestine and liver, which reduces the amount of drug delivered to the systemic circulation, is termed *presystemic elimination*, *presystemic extraction*, or *first-pass elimination*.

FIGURE 67-3

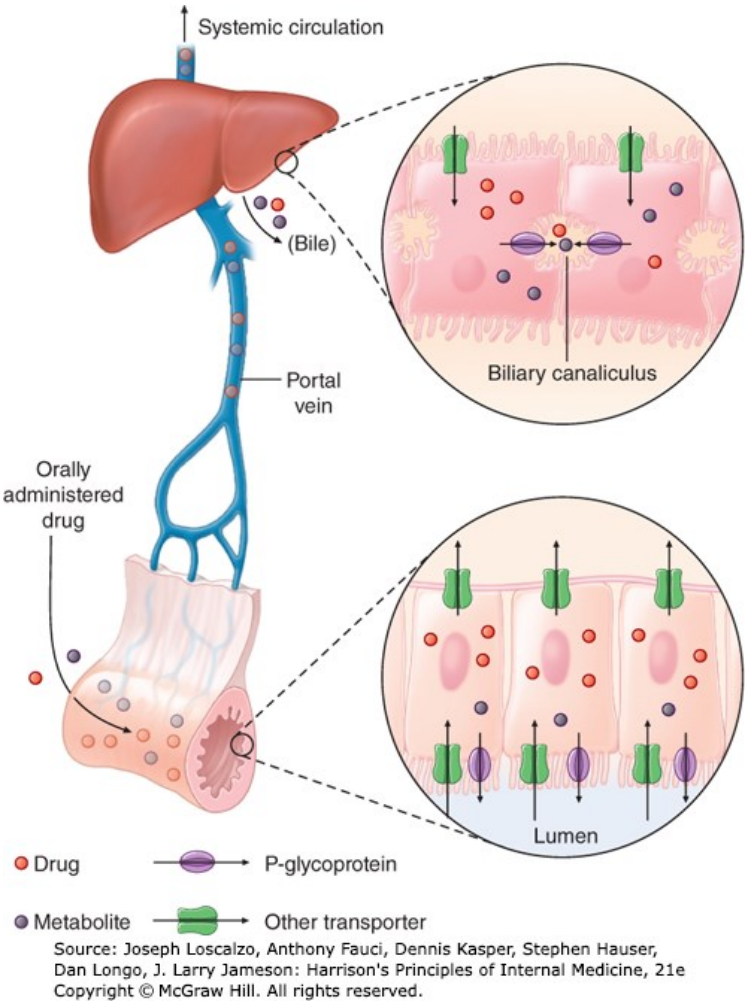
Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.)

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Mechanism of presystemic elimination. After drug enters the enterocyte, it can undergo metabolism, excretion into the intestinal lumen, or transport into the portal vein. Similarly, the hepatocyte may accomplish metabolism and biliary excretion prior to the entry of drug and metabolites to the systemic circulation. *(Adapted by permission from DM Roden, in DP Zipes, J Jalife [eds]: Cardiac Electrophysiology: From Cell to Bedside, 4th ed. Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

DRUG TRANSPORT

Drug movement across the membrane of any cell, including enterocytes and hepatocytes, is a combination of passive diffusion and active transport, mediated by specific drug uptake and efflux molecules. One widely studied drug transport molecule is the drug efflux pump P-glycoprotein, the product of the *ABCB1* (or *MDR1*) gene. P-glycoprotein is expressed on the apical aspect of the enterocyte and on the canalicular aspect of the hepatocyte (Fig. 67-3). In both locations, it serves as an efflux pump, limiting availability of drug to the systemic circulation. P-glycoprotein–mediated drug efflux from cerebral capillaries limits drug brain penetration and is an important component of the blood-brain barrier. Other transporters mediate uptake into cells of drugs and endogenous substrates such as vitamins or nutrients.

DRUG METABOLISM

Drug metabolism generates compounds that are usually more polar and, hence, more readily excreted than parent drug. Metabolism takes place predominantly in the liver but can occur at other sites such as kidney, intestinal epithelium, lung, and plasma. Phase I metabolism involves chemical modification, most often oxidation accomplished by members of the cytochrome P450 (CYP) monooxygenase superfamily. CYPs and other molecules that are especially important for drug metabolism are presented in **Table 67-1**, and each drug may be a substrate for one or more of these enzymes. Phase II metabolism involves conjugation of specific endogenous compounds to drugs or their metabolites. The enzymes that accomplish phase II reactions include glucuronyl-, acetyl-, sulfo-, and methyltransferases. Drug metabolites may exert important pharmacologic activity, as discussed

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 67-1

Molecular Pathways Mediating Drug Disposition

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Phase II metabolism involves conjugation of specific endogenous compounds to drugs or their metabolites. The enzymes that accomplish phase II reactions include glucuronyl-, acetyl-, sulfo-, and methyltransferases. Drug metabolites may exert important pharmacologic activity, as discussed further below. Therapeutic antibodies are very slowly eliminated (allowing infrequent dosing, e.g., monthly injections), probably by lysosomal uptake and degradation.

TABLE 67-1

Molecular Pathways Mediating Drug Disposition

ENZYME	SUBSTRATES[a]	INHIBITORS[a]
CYP3A	Calcium channel blockers	Amiodarone
	Antiarrhythmics (lidocaine, quinidine, mexiletine)	Ketoconazole, itraconazole
	HMG-CoA reductase inhibitors ("statins"; see text)	Erythromycin, clarithromycin
	Cyclosporine, tacrolimus	Ritonavir
	Indinavir, saquinavir, ritonavir	Gemfibrozil and other fibrates
CYP2D6[b]	Timolol, metoprolol, carvedilol	Quinidine (even at ultra-low doses)
	Propafenone, flecainide	Tricyclic antidepressants
	Tricyclic antidepressants	Fluoxetine, paroxetine
	Fluoxetine, paroxetine	
CYP2C9[b]	Warfarin	Amiodarone
	Phenytoin	Fluconazole
	Glipizide	Phenytoin
	Losartan	
CYP2C19[b]	Omeprazole	Omeprazole
	Mephenytoin	
	Clopidogrel	
CYP2B6[b]	Efavirenz	
Thiopurine S-methyltransferase[b]	6-Mercaptopurine, azathioprine	
N-acetyltransferase[b]	Isoniazid	
	Procainamide	
	Hydralazine	
Pseudocholinesterase[b]	Succinylcholine	

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

	Some sulfonamides	
UGT1A1[b]	Irinotecan	
Pseudocholinesterase[b]	Succinylcholine	
TRANSPORTER	**SUBSTRATES[a]**	**INHIBITORS[a]**
P-glycoprotein	Digoxin	Quinidine
	HIV protease inhibitors	Amiodarone
	Many CYP3A substrates	Verapamil
		Cyclosporine
		Itraconazole
		Erythromycin
SLCO1B1[b]	Simvastatin and some other statins	

[a]Inhibitors affect the molecular pathway and thus may decrease substrate metabolism. [b]Clinically important genetic variants described; **see Chap. 68**.

Note: A listing of CYP substrates, inhibitors, and inducers is maintained at *https://drug-interactions.medicine.iu.edu/MainTable.aspx*.

Clinical Implications of Altered Bioavailability

Some drugs undergo near-complete presystemic metabolism and thus cannot be administered orally. Nitroglycerin cannot be used orally because it is completely extracted prior to reaching the systemic circulation. The drug is, therefore, used by the sublingual, transdermal, or intravascular routes, which bypass presystemic metabolism.

Some drugs with very extensive presystemic metabolism can still be administered by the oral route, using much higher doses than those required intravenously. Thus, a typical intravenous dose of verapamil is 1–5 mg, compared to a usual single oral dose of 40–120 mg. Administration of low-dose aspirin can result in exposure of cyclooxygenase in platelets in the portal vein to the drug, but systemic sparing because of first-pass aspirin deacylation in the liver. This is an example of presystemic metabolism being exploited to therapeutic advantage.

PLASMA HALF-LIFE

Most pharmacokinetic processes, such as elimination, are first-order; that is, the rate of the process depends on the amount of drug present. Elimination can occasionally be zero-order (fixed amount eliminated per unit time), and this can be clinically important (see "Principles of Dose Selection," later in this chapter). In the simplest pharmacokinetic model (Fig. 67-2A), a drug bolus (D) is administered instantaneously to a central compartment, from which drug elimination occurs as a first-order process. Occasionally, central and other compartments correspond to physiologic spaces (e.g., plasma volume), whereas in other cases, they are simply mathematical functions used to describe drug disposition. The first-order nature of drug elimination leads directly to the relationship describing drug concentration (C) at any time (t) following the bolus:

$$C = \frac{D}{V_c} \cdot e^{(-0.69t/t_{1/2})}$$

where V_c is the volume of the compartment into which drug is delivered and $t_{1/2}$ is elimination half-life. As a consequence of this relationship, a plot of

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

In some cases, drug is removed from the central compartment not only by elimination but also by distribution into peripheral compartments. In this

where V_c is the volume of the compartment into which drug is delivered and $t_{1/2}$ is elimination half-life. As a consequence of this relationship, a plot of the logarithm of concentration versus time is a straight line (Fig. 67-2A, inset). *Half-life* is the time required for 50% of a first-order process to be completed. Thus, 50% of drug elimination is achieved after one drug-elimination half-life, 75% after two, 87.5% after three, etc. In practice, first-order processes such as elimination are near-complete after four to five half-lives.

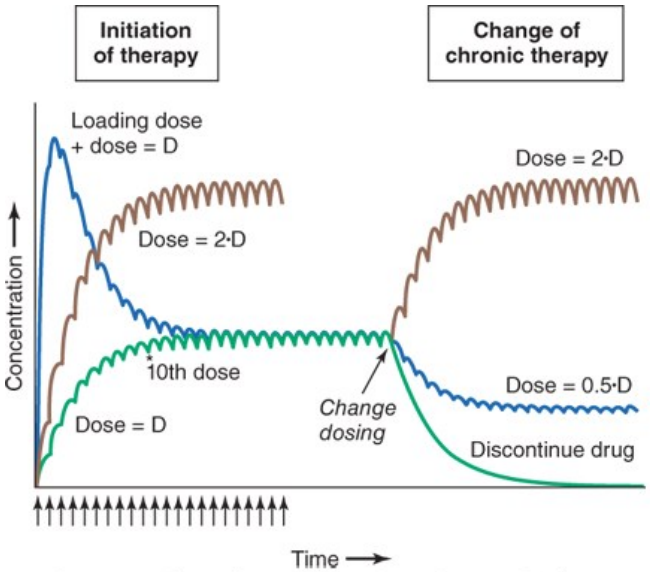
In some cases, drug is removed from the central compartment not only by elimination but also by distribution into peripheral compartments. In this case, the plot of plasma concentration versus time after a bolus may demonstrate two (or more) exponential components (Fig. 67-2B). In general, the initial rapid drop in drug concentration represents not elimination but drug distribution into and out of peripheral tissues (also first-order processes), while the slower component represents drug elimination; the initial precipitous decline is usually evident with administration by intravenous but not by other routes. Drug concentrations at peripheral sites are determined by a balance between drug distribution to and redistribution from those sites, as well as by elimination. Once distribution is near-complete (four to five distribution half-lives), plasma and tissue concentrations decline in parallel.

Clinical Implications of Half-Life Measurements

The elimination half-life not only determines the time required for drug concentrations to fall to near-immeasurable levels after a single bolus, it is also the sole determinant of the time required for steady-state plasma concentrations to be achieved after any change in drug dosing **(Fig. 67-4)**. This applies to the initiation of chronic drug therapy (whether by multiple oral doses or by continuous intravenous infusion), a change in chronic drug dose or dosing interval, or discontinuation of drug.

FIGURE 67-4

Drug accumulation to steady state. In this simulation, drug was administered (arrows) at intervals = 50% of the elimination half-life. Steady state is achieved during initiation of therapy after ~5 elimination half-lives, or 10 doses. A loading dose did not alter the eventual steady state achieved. A doubling of the dose resulted in a doubling of the steady state but the same time course of accumulation. Once steady state is achieved, a change in dose (increase, decrease, or drug discontinuation) results in a new steady state in ~5 elimination half-lives. *(Adapted by permission from DM Roden, in DP Zipes, J Jalife [eds]: Cardiac Electrophysiology: From Cell to Bedside, 4th ed. Philadelphia, Saunders, 2003. Copyright 2003 with permission from Elsevier.)*



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Steady state describes the situation during chronic drug administration when the amount of drug administered per unit time equals drug eliminated per unit time. With a continuous intravenous infusion, plasma concentrations at steady state are stable, while with chronic oral drug administration, plasma concentrations vary during the dosing interval, but the time-concentration profile between dosing intervals is stable (Fig. 67-4).

DRUG DISTRIBUTION

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

distribution of digoxin and tricyclic antidepressants is hundreds of liters, obviously exceeding total-body volume. Such drugs are not readily removed by dialysis, an important consideration in overdose.

In a typical 70-kg human, plasma volume is ~3 L, blood volume is ~5.5 L, and extracellular water outside the vasculature is ~20 L. The volume of distribution of drugs extensively bound to plasma proteins but not to tissue components approaches plasma volume; warfarin is an example. By contrast, for drugs highly bound to tissues, the volume of distribution can be far greater than any physiologic space. For example, the volume of distribution of digoxin and tricyclic antidepressants is hundreds of liters, obviously exceeding total-body volume. Such drugs are not readily removed by dialysis, an important consideration in overdose.

Clinical Implications of Drug Distribution

In some cases, pharmacologic effects require drug distribution to peripheral sites. In this instance, the time course of drug delivery to and removal from these sites determines the time course of drug effects; anesthetic uptake into the central nervous system (CNS) is an example.

LOADING DOSES

For some drugs, the indication may be so urgent that administration of "loading" dosages is required to achieve rapid elevations of drug concentration and therapeutic effects earlier than with chronic maintenance therapy (Fig. 67-4). Nevertheless, the time required for a true steady state to be achieved is still determined only by the elimination half-life.

RATE OF INTRAVENOUS DRUG ADMINISTRATION

Although the simulations in Fig. 67-2 use a single intravenous bolus, this is usually inappropriate in practice because side effects related to transiently very high concentrations can result. Rather, drugs are more usually administered orally or as a slower intravenous infusion. Some drugs are so predictably lethal when infused too rapidly that special precautions should be taken to prevent accidental boluses. For example, solutions of potassium for intravenous administration >20 mEq/L should be avoided in all but the most exceptional and carefully monitored circumstances. This minimizes the possibility of cardiac arrest due to accidental increases in infusion rates of more concentrated solutions.

Transiently high drug concentrations after rapid intravenous administration can occasionally be used to advantage. The use of midazolam for intravenous sedation, for example, depends upon its rapid uptake by the brain during the distribution phase to produce sedation quickly, with subsequent egress from the brain during the redistribution of the drug as equilibrium is achieved.

Similarly, adenosine must be administered as a rapid bolus in the treatment of reentrant supraventricular tachycardias **(Chap. 246)** to prevent elimination by very rapid ($t_{1/2}$ of seconds) uptake into erythrocytes and endothelial cells before the drug can reach its clinical site of action, the atrioventricular node.

Clinical Implications of Altered Protein Binding

Many drugs circulate in the plasma partly bound to plasma proteins. Since only unbound (free) drug can distribute to sites of pharmacologic action, drug response is related to the free rather than the total circulating plasma drug concentration. In chronic kidney or liver disease, protein binding may be decreased and thus drug actions increased. In some situations (myocardial infarction, infection, surgery), acute phase reactants transiently increase binding of some drugs and thus decrease efficacy. These changes assume the greatest clinical importance for drugs that are highly protein-bound since even a small change in protein binding can result in large changes in free drug; for example, a decrease in binding from 99 to 98% doubles the free drug concentration from 1 to 2%. For some drugs (e.g., phenytoin), monitoring free rather than total drug concentrations can be useful.

DRUG ELIMINATION

Drug elimination reduces the amount of drug in the body over time. An important approach to quantifying this reduction is to consider that drug concentrations at the beginning and end of a time period are unchanged, and that a specific volume of the body has been "cleared" of the drug during that time period. This defines clearance as volume/time. Clearance includes both drug metabolism and excretion.

Clinical Implications of Altered Clearance

While elimination half-life determines the time required to achieve steady-state plasma concentration (C_{ss}), the *magnitude* of that steady state is determined by clearance (*Cl*) and dose alone. For a drug administered as an intravenous infusion, this relationship is:

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

When a drug is administered orally, the average plasma concentration within a dosing interval ($C_{avg,ss}$) replaces C_{ss}, and the dosage (dose per unit

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

that time period. This defines clearance as volume/time. Clearance includes both drug metabolism and excretion.

While elimination half-life determines the time required to achieve steady-state plasma concentration (C_{ss}), the *magnitude* of that steady state is determined by clearance (*Cl*) and dose alone. For a drug administered as an intravenous infusion, this relationship is:

C_{ss} = dosing rate/*Cl* or dosing rate = *Cl* • C_{ss}

When a drug is administered orally, the average plasma concentration within a dosing interval ($C_{avg,ss}$) replaces C_{ss}, and the dosage (dose per unit time) must be increased if bioavailability (*F*) is <100%:

Dose/time = *Cl* • $C_{avg,ss}$/*F*

Genetic variants, drug interactions, or diseases that reduce the activity of drug-metabolizing enzymes or excretory mechanisms lead to decreased clearance and, hence, a requirement for a downward dose adjustment to avoid toxicity. Conversely, some drug interactions and genetic variants increase the function of drug elimination pathways, and hence, increased drug dosage is necessary to maintain a therapeutic effect.

ACTIVE DRUG METABOLITES

Metabolites may produce effects similar to, overlapping with, or distinct from those of the parent drug. Accumulation of the major metabolite of procainamide, *N*-acetylprocainamide (NAPA), likely accounts for marked QT prolongation and torsades de pointes ventricular tachycardia **(Chap. 252)** during therapy with procainamide. Neurotoxicity during therapy with the opioid analgesic meperidine is likely due to accumulation of normeperidine, especially in renal disease.

Prodrugs are inactive compounds that require metabolism to generate active metabolites that mediate the drug effects. Examples include many angiotensin-converting enzyme (ACE) inhibitors, the angiotensin receptor blocker losartan, the antineoplastic irinotecan, the antiestrogen tamoxifen, the analgesic codeine (whose active metabolite morphine probably underlies the opioid effect during codeine administration), and the antiplatelet drug clopidogrel. Drug metabolism has also been implicated in bioactivation of procarcinogens and in the generation of reactive metabolites that mediate certain ADRs (e.g., acetaminophen hepatotoxicity, discussed below).

THE CONCEPT OF HIGH-RISK PHARMACOKINETICS

When plasma concentrations of active drug depend exclusively on a single metabolic pathway, any condition that inhibits that pathway (be it disease related, genetic, or due to a drug interaction) can lead to dramatic changes in drug concentrations and marked variability in drug action. Two mechanisms can generate highly variable drug concentrations and effects through such "high-risk pharmacokinetics." *First*, variability in bioactivation of a prodrug can lead to striking variability in drug action; examples include decreased CYP2D6 activity, which prevents analgesia by codeine, and decreased CYP2C19 activity, which reduces the antiplatelet effects of clopidogrel. The s*econd* setting is drug elimination that relies on a single pathway. In this case, inhibition of the elimination pathway by genetic variants or by administration of inhibiting drugs leads to marked elevation of drug concentration and, for drugs with a narrow therapeutic window, an increased likelihood of dose-related toxicity. The active S-enantiomer of the anticoagulant warfarin is eliminated by CYP2C9, and co-administration of amiodarone or phenytoin, CYP2C9 inhibitors, may therefore increase the risk of bleeding unless the dose is decreased. When drugs undergo elimination by multiple-drug metabolizing or excretory pathways, absence of one pathway (due to a genetic variant or drug interaction) is much less likely to have a large impact on drug concentrations or drug actions.

PRINCIPLES OF PHARMACODYNAMICS

Time Course of Drug Action

Pharmacokinetic parameters, such as half-life and clearance, explain drug concentrations over time, but understanding the action of a drug over time (pharmacodynamics) often requires an understanding of its precise mechanism of action. Drugs act through interactions with drug targets, often in specific tissues, and with a cascade of downstream consequences. For drugs used in the urgent treatment of acute symptoms, little or no delay is anticipated (or desired) between the administration of the drug, the drug-target interaction, and the development of a clinical effect. Examples of such acute situations include vascular thrombosis, shock, or status epilepticus.

For many conditions, however, the indication for therapy is less urgent, and a delay in the onset of action clinically acceptable. Delay can be due to pharmacokinetic mechanisms such as slow elimination (resulting in slow accumulation to steady state), slow uptake into the target tissue, or slow

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

time frame, as evidenced by the biological half-life of 24–36 h. Other examples include proton pump inhibitors, which irreversibly bind the hydrogen/potassium adenosine triphosphatase enzyme and thus affect acid secretion for the lifetime of that enzyme, and the irreversible antiplatelet

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

For many conditions, however, the indication for therapy is less urgent, and a delay in the onset of action clinically acceptable. Delay can be due to pharmacokinetic mechanisms such as slow elimination (resulting in slow accumulation to steady state), slow uptake into the target tissue, or slow accumulation of active metabolites. A common pharmacodynamic explanation for such a delay is the biological mechanism of action. For example, the glucocorticoid prednisolone has a plasma half-life of about 60 min. The mechanism of action, however, involves binding of the glucocorticoid receptor, translocation to the cell nucleus, and alterations in gene transcription. These downstream effects alter immune function for a much longer time frame, as evidenced by the biological half-life of 24–36 h. Other examples include proton pump inhibitors, which irreversibly bind the hydrogen/potassium adenosine triphosphatase enzyme and thus affect acid secretion for the lifetime of that enzyme, and the irreversible antiplatelet drugs, which exert effects for the duration of the life of the platelet.

Drug Effects May Be Disease Specific

A drug may produce no action or a different spectrum of actions in unaffected individuals compared to patients with underlying disease. Further, concomitant disease can complicate interpretation of response to drug therapy, especially ADRs. For example, high doses of anticonvulsants such as phenytoin may cause neurologic symptoms, which may be confused with the underlying neurologic disease. Similarly, increasing dyspnea in a patient with chronic lung disease receiving amiodarone therapy could be due to the drug, underlying disease, or an intercurrent cardiopulmonary problem. As a result, alternate antiarrhythmic therapies may be preferable in patients with chronic lung disease.

While drugs interact with specific molecular receptors, drug effects may vary over time, even if stable drug and metabolite concentrations are maintained. The drug-receptor interaction occurs in a complex biologic milieu that can vary to modulate the drug effect. For example, ion channel blockade by drugs, an important anticonvulsant and antiarrhythmic effect, is often modulated by membrane potential, itself a function of factors such as extracellular potassium or local ischemia. Receptors may be up- or downregulated by disease or by the drug itself. For example, β-adrenergic blockers upregulate β-receptor density during chronic therapy. While this effect does not usually result in resistance to the therapeutic effect of the drugs, it may produce severe agonist-mediated effects (such as hypertension or tachycardia) if the blocking drug is abruptly withdrawn.

As molecular mechanisms of disease become better defined, drugs targeting those mechanisms are being introduced into practice. Antineoplastic agents targeting mutant kinases overexpressed in cancers (e.g., *BRAF* V600E in melanoma, hairy cell leukemia, and other malignancies) are revolutionizing cancer care. Ivacaftor was originally developed and marketed for patients with cystic fibrosis (CF) carrying the G551D mutation in the disease gene *CFTR* **(Chap. 291)**. While the most common *CFTR* mutations causing CF generate normal chloride channels that are not correctly trafficked to the cell surface, G551D channels are trafficked normally but do not conduct chloride correctly, and ivacaftor corrects this “gating” defect. Following initial marketing for only G551D patients (5% of all CF patients), the U.S. Food and Drug Administration (FDA) approved ivacaftor for use in patients carrying other *CFTR* mutations that confer gating defects corrected by ivacaftor in vitro.

PRINCIPLES OF DOSE SELECTION

The desired goal of therapy with any drug is to maximize the likelihood of a beneficial effect while minimizing the risk of ADRs. Previous experience with the drug, in controlled clinical trials or in postmarketing use, defines the relationships between dose or plasma concentration and these dual effects (Fig. 67-1) and has important implications for initiation of drug therapy:

1. *The target drug effect should be defined when drug treatment is started.* With some drugs, the desired effect may be difficult to measure objectively, or the onset of efficacy can be delayed for weeks or months; drugs used in the treatment of cancer and psychiatric disease are examples. Sometimes a drug is used to treat a symptom, such as pain or palpitations, and here it is the patient who will report whether the selected dose is effective. In yet other settings, such as anticoagulation or hypertension, the desired response can be repeatedly and objectively assessed by simple clinical or laboratory tests.

2. *The nature of anticipated toxicity often dictates the starting dose.* If side effects are minor, it may be acceptable to start chronic therapy at a dose highly likely to achieve efficacy and down-titrate if side effects occur. However, this approach is rarely, if ever, justified if the anticipated toxicity is serious or life-threatening; in this circumstance, it is more appropriate to initiate therapy with the lowest dose that may produce a desired effect. In cancer chemotherapy, it is common practice to use maximally tolerated doses.

3. *The above considerations do not apply if these relationships between dose and effects cannot be defined.* This is especially relevant to some ADRs (discussed further below) whose development is not readily related to drug dose.

4. *If a drug dose does not achieve its desired effect, a dosage increase is justified only if toxicity is absent and the likelihood of serious toxicity is small.*

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

expected. A complete therapeutic response is often absent with antihypertensive or antidepressant drugs, and a major challenge in contemporary therapeutics is to identify patient-specific predictors of response to individual drugs. Other explanations for failure of efficacy include drug

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

4. *If a drug dose does not achieve its desired effect, a dosage increase is justified only if toxicity is absent and the likelihood of serious toxicity is small.*

Failure of Efficacy

Even assuming the diagnosis is correct and the correct drug and dose are prescribed, drugs may fail to be effective because 100% efficacy is not expected. A complete therapeutic response is often absent with antihypertensive or antidepressant drugs, and a major challenge in contemporary therapeutics is to identify patient-specific predictors of response to individual drugs. Other explanations for failure of efficacy include drug interactions, noncompliance, or unexpectedly low drug concentration due to administration of expired or degraded drug. These are situations in which measurement of plasma drug concentrations, if available, can be especially useful. Noncompliance is an especially frequent problem in the long-term treatment of diseases such as hypertension and epilepsy, occurring in ≥25% of patients in therapeutic environments in which no special effort is made to involve patients in the responsibility for their own health. Multidrug regimens with multiple doses per day are especially prone to noncompliance.

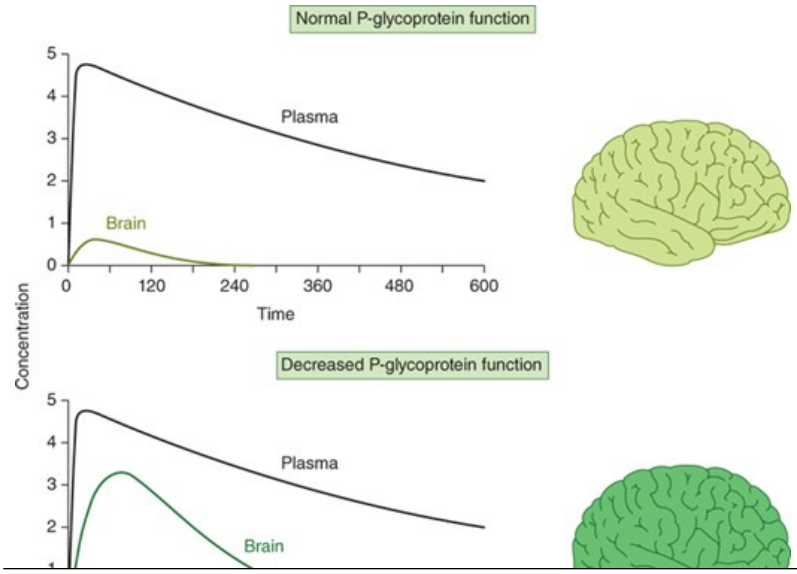
Monitoring response to therapy, by physiologic measures or by plasma concentration measurements, requires an understanding of the relationships between plasma concentration and anticipated effects. For example, measurement of QT interval is used during treatment with sotalol or dofetilide to avoid marked QT prolongation that can herald serious arrhythmias. In this setting, evaluating the electrocardiogram at the time of anticipated peak plasma concentration and effect (e.g., 1–2 h postdose at steady state) is most appropriate. Maintained high vancomycin levels carry a risk of nephrotoxicity, so dosages should be adjusted on the basis of plasma concentrations measured at trough (predose). Similarly, for dose adjustment of other drugs (e.g., anticonvulsants), concentration should be measured at its lowest during the dosing interval, just prior to a dose at steady state (Fig. 67-4), to ensure a maintained therapeutic effect.

Concentration of Drugs in Plasma as a Guide to Therapy

Factors such as interactions with other drugs, disease-induced alterations in elimination and distribution, and genetic variation in drug disposition combine to yield a wide range of plasma levels in patients given the same dose. Hence, if a predictable relationship can be established between plasma drug concentration and beneficial or adverse drug effect, measurement of plasma levels can provide a valuable tool to guide selection of an optimal dose, especially when there is a narrow range between the plasma levels yielding therapeutic and adverse effects. Such therapeutic drug monitoring is commonly used with certain types of drugs including many anticonvulsants, antirejection agents, antiarrhythmics, and antibiotics. By contrast, if no such relationship can be established (e.g., if drug access to important sites of action outside plasma is highly variable), monitoring plasma concentration may not provide an accurate guide to therapy **(Fig. 67-5)**.

FIGURE 67-5

Drug concentrations in specific tissues may not always parallel those in plasma. For example, the efflux pump P-glycoprotein excludes drugs from the endothelium of capillaries in the brain and so constitutes a key element of the blood-brain barrier. Reduced P-glycoprotein function (e.g., due to drug interactions) can thus increase penetration of substrate drugs into the brain, even when plasma concentrations are unchanged.

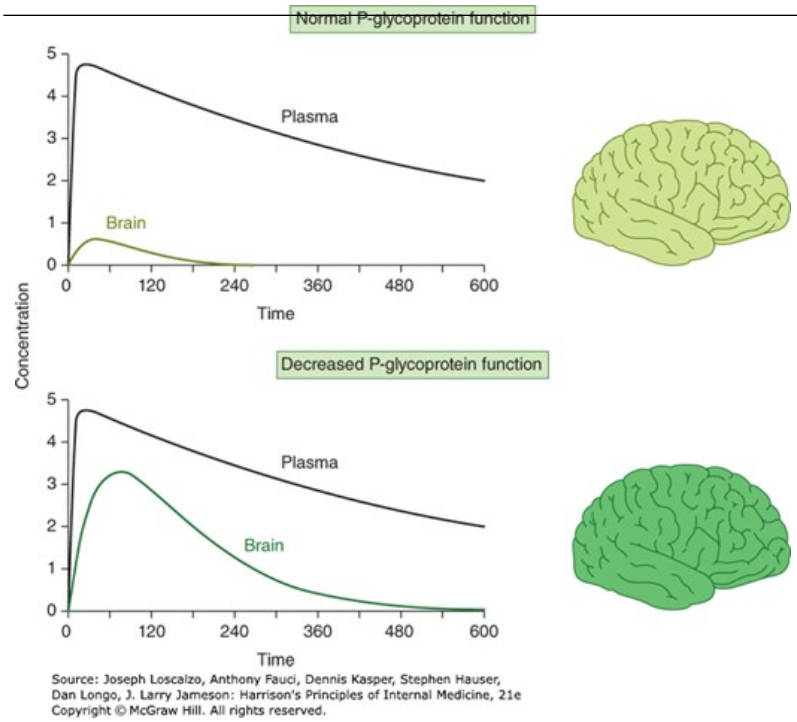


Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The common situation of first-order elimination implies that average, maximum, and minimum steady-state concentrations are related linearly to the

Access Provided by:
SILVERCHAIR



The common situation of first-order elimination implies that average, maximum, and minimum steady-state concentrations are related linearly to the dosing rate. Accordingly, the maintenance dose may be adjusted on the basis of the ratio between the desired and measured concentrations *at steady state*; for example, if a doubling of the steady-state plasma concentration is desired, the dose should be doubled. This does not apply to drugs eliminated by zero-order kinetics (fixed amount per unit time), where small dosage increases will produce disproportionate increases in plasma concentration; examples include phenytoin and theophylline.

If an increase in dosage is needed, this is usually best achieved by increasing the drug dose and leaving the dosing interval constant (e.g., by giving 200 mg every 8 h instead of 100 mg every 8 h). However, this approach is acceptable only if the resulting maximum concentration is not toxic and the trough value does not fall below the minimum effective concentration for an undesirable period of time. Alternatively, the steady state may be changed by altering the frequency of intermittent dosing but not the size of each dose. In this case, the magnitude of the fluctuations around the average steady-state level will change—the shorter the dosing interval, the smaller the difference between peak and trough levels.

EFFECTS OF DISEASE ON DRUG CONCENTRATION AND RESPONSE

RENAL DISEASE

Renal excretion of parent drug and metabolites is generally accomplished by glomerular filtration and by specific drug transporters. If a drug or its metabolites are primarily excreted through the kidneys and increased drug levels are associated with ADRs (an example of “high-risk pharmacokinetics” described above), drug dosages must be reduced in patients with renal dysfunction to avoid toxicity. The antiarrhythmics dofetilide and sotalol undergo predominant renal excretion and carry a risk of QT prolongation and arrhythmias if doses are not reduced in renal disease. In end-stage renal disease, sotalol has been given as 40 mg after dialysis (every second day), compared to the usual daily dose, 80–120 mg every 12 h. At approved doses, the anticoagulant edoxaban appears to be somewhat more effective in subjects with mild renal dysfunction, possibly reflecting higher drug levels. The narcotic analgesic meperidine undergoes extensive hepatic metabolism, so that renal failure has little effect on its plasma concentration. However, its metabolite, normeperidine, does undergo renal excretion, accumulates in renal failure, and probably accounts for the signs of CNS excitation, such as irritability, twitching, and seizures, that appear when multiple doses of meperidine are administered to patients with renal disease. Protein binding of some drugs (e.g., phenytoin) may be altered in uremia, so measuring free drug concentration may be desirable.

In non-end-stage renal disease, changes in renal drug clearance are generally proportional to those in creatinine clearance, which may be measured directly or estimated from the serum creatinine. This estimate, coupled with the knowledge of how much drug is normally excreted renally versus nonrenally, allows an estimate of the dose adjustment required. In practice, most decisions involving dosing adjustment in patients with renal failure use published recommended adjustments in dosage or dosing interval based on the severity of renal dysfunction indicated by creatinine clearance

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

LIVER DISEASE

ACCESS Medicine
Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

nonrenally, allows an estimate of the dose adjustment required. In practice, most decisions involving dosing adjustment in patients with renal failure use published recommended adjustments in dosage or dosing interval based on the severity of renal dysfunction indicated by creatinine clearance. Any such modification of dose is a first approximation and should be followed by plasma concentration data (if available) and clinical observation to further optimize therapy for the individual patient.

LIVER DISEASE

Standard tests of liver function are not useful in adjusting doses in diseases like hepatitis or cirrhosis. First-pass metabolism may decrease, leading to increased oral bioavailability as a consequence of disrupted hepatocyte function, altered liver architecture, and portacaval shunts. The oral bioavailability for high first-pass drugs such as morphine, meperidine, midazolam, and nifedipine is almost doubled in patients with cirrhosis, compared to those with normal liver function. Therefore, the size of the oral dose of such drugs should be reduced in this setting.

HEART FAILURE AND SHOCK

Under conditions of decreased tissue perfusion, the cardiac output is redistributed to preserve blood flow to the heart and brain at the expense of other tissues **(Chap. 257)**. As a result, drugs may be distributed into a smaller volume of distribution, higher drug concentrations will be present in the plasma, and the tissues that are best perfused (the brain and heart) will be exposed to these higher concentrations, resulting in increased CNS or cardiac effects. In addition, decreased perfusion of the kidney and liver may impair drug clearance. Another consequence of severe heart failure is decreased gut perfusion, which may reduce drug absorption and thus lead to reduced or absent effects of orally administered therapies.

DRUG USE IN THE ELDERLY

In the elderly, multiple pathologies and medications used to treat them result in more drug interactions and ADRs. Aging also results in changes in organ function, especially of the organs involved in drug disposition. Initial doses should be less than the usual adult dosage and should be increased slowly. The number of medications, and doses per day, should be kept as low as possible.

Even in the absence of kidney disease, renal clearance may be reduced by 35–50% in elderly patients. Dosages should be adjusted on the basis of creatinine clearance. Aging also results in a decrease in the size of, and blood flow to, the liver and possibly in the activity of hepatic drug-metabolizing enzymes; accordingly, the hepatic clearance of some drugs is impaired in the elderly. As with liver disease (above), these changes are not readily predicted.

Elderly patients may display altered drug sensitivity. Examples include increased analgesic effects of opioids, increased sedation from benzodiazepines and other CNS depressants, and increased risk of bleeding while receiving anticoagulant therapy, even when clotting parameters are well controlled. Exaggerated responses to cardiovascular drugs are also common because of the impaired responsiveness of normal homeostatic mechanisms. Conversely, the elderly display decreased sensitivity to β-adrenergic receptor blockers.

ADRs are especially common in the elderly because of altered pharmacokinetics and pharmacodynamics, the frequent use of multidrug regimens, and concomitant disease. For example, use of long half-life benzodiazepines is linked to the occurrence of hip fractures in elderly patients, perhaps reflecting both a risk of falls from these drugs (due to increased sedation) and the increased incidence of osteoporosis in elderly patients. In population surveys of the noninstitutionalized elderly, as many as 10% had at least one ADR in the previous year.

DRUG USE IN CHILDREN

Although there are very few pediatric-specific drugs, there are many pediatric-specific drug indications (e.g., intravenous immunoglobulin and aspirin for Kawasaki disease) and ADRs (e.g., pyloric stenosis after erythromycin exposure in infants). Drug metabolism and drug response pathways mature at different rates after birth, and the relative size of various body compartments and function of various organs change during development. There is increased motivation to avoid organ toxicity, given the anticipated long post-drug-exposure life expectancy. There are few studies providing empiric evidence to guide pediatric dosing. In practice, doses are adjusted for size (weight or body surface area) as a first approximation unless age-specific data are available. As in adults, the lowest doses anticipated to achieve clinical benefit are generally prescribed, potentially followed by titration.

INTERACTIONS BETWEEN DRUGS

Drug interactions can complicate therapy by increasing or decreasing the action of a drug; interactions may be based on changes in drug disposition or in drug response in the absence of changes in drug levels **(Table 67-3)**. *Interactions must be considered in the differential diagnosis of any unusual*

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

list all medications, including agents not often volunteered during questioning, such as OTC drugs, health food supplements, and topical agents such as eye drops. Lists of interactions are available from a number of electronic sources. While it is unrealistic to expect the practicing physician to

ACCESS Medicine®

Access Provided by: SILVERCHAIR INFORMATION/SYSTEMS

Drug interactions can complicate therapy by increasing or decreasing the action of a drug; interactions may be based on changes in drug disposition or in drug response in the absence of changes in drug levels **(Table 67-2)**. *Interactions must be considered in the differential diagnosis of any unusual response occurring during drug therapy.* Prescribers should recognize that patients often come to them with a legacy of drugs acquired during previous medical experiences, often with multiple physicians who may not be aware of all the patient's medications. A meticulous drug history should list all medications, including agents not often volunteered during questioning, such as OTC drugs, health food supplements, and topical agents such as eye drops. Lists of interactions are available from a number of electronic sources. While it is unrealistic to expect the practicing physician to memorize these, certain drugs consistently run the risk of generating interactions, often by inhibiting or inducing specific drug elimination pathways; these include CYP2D6, CYP3A, and P-glycoprotein inhibitors (Table 67-1) and CYP3A/P-glycoprotein inducers (Table 67-2). Accordingly, when these drugs are started or stopped, prescribers must be especially alert to the possibility of interactions.

TABLE 67-2

Drug Interactions

MECHANISM	EXAMPLE
Pharmacokinetic Interactions Causing Decreased Drug Effect	
Decreased absorption due to drug binding in the gut	Antacids or bile acid sequestrants decrease the absorption of many drugs: Antacids/tetracyclines Cholestyramine/digoxin
Decreased solubility due to altered gastric pH	H_2 receptor blockers or proton pump inhibitors decrease solubility and absorption of weak bases: Omeprazole/ketoconazole
Induction of drug metabolism and/or drug transport: Rifampin Carbamazepine Phenytoin St. John's wort Glutethimide (also smoking, exposure to chlorinated insecticides, and chronic alcohol ingestion)	Decreased concentrations and effects of: Warfarin Quinidine Cyclosporine Losartan Oral contraceptives Methadone Dabigatran
Decreased prodrug bioactivation	Proton pump inhibitors may prevent clopidogrel bioactivation CYP2D6 inhibitors (fluoxetine, paroxetine, quinidine, and others) may prevent codeine bioactivation
Reduced delivery of drug to active sites of action	Tricyclics prevent clonidine uptake into adrenergic neurons, preventing antihypertensive effects
Pharmacokinetic Interactions Causing Increased Drug Effect	
Inhibited drug metabolism	Cimetidine (inhibits many CYPs): Warfarin Theophylline Phenytoin CYP2D6 inhibitors[a]/β blockers CYP3A inhibitors[a]: HMG-CoA reductase inhibitors Colchicine (toxicity risk) Decreased cyclosporine dose requirement
	Digoxin Dabigatran

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

	Colchicine (toxicity risk) Decreased cyclosporine dose requirement
Inhibited drug transport	Amiodarone (inhibits many CYPs and P-glycoprotein): Warfarin Digoxin Dabigatran
Inhibition of drug metabolism causing accumulation of toxic metabolites	Allopurinol (xanthine oxidase inhibitor) inhibits an alternate pathway for azathioprine and 6-mercaptopurine elimination, increasing risk for toxicity
Decreased elimination due to altered renal function	Inhibitors of renal tubular transport (phenylbutazone, probenecid, salicylates) increase methotrexate toxicity
Pharmacodynamic Drug Interactions	
Combined effects on the same biologic process	Excess bleeding with combinations of antiplatelet drugs, anticoagulants, and NSAIDs Long QT–related arrhythmias with QT-prolonging antiarrhythmics plus diuretics Hyperkalemia with ACE inhibitors plus potassium Hypotension with nitrates plus sildenafil
Antagonistic effects on the same biologic process	Loss of antihypertensive drug effects with NSAIDs

[a]See Table 67-1.

Abbreviations: ACE, angiotensin-converting enzyme; CYP, cytochrome P; NSAID, nonsteroidal anti-inflammatory drug.

ADVERSE DRUG REACTIONS

The beneficial effects of drugs are coupled with the inescapable risk of untoward effects. The morbidity and mortality from these ADRs often present diagnostic problems because they can involve every organ and system of the body and may be mistaken for signs of underlying disease. In addition, some surveys have suggested that drug therapy for a range of chronic conditions such as psychiatric disease or hypertension does not achieve its desired goal in up to half of treated patients; thus, the most common “adverse” drug effect may be failure of efficacy.

ADRs can be classified in two broad groups. Type A reactions result from exaggeration of an intended pharmacologic action of the drug, such as increased bleeding with anticoagulants or bone marrow suppression with some antineoplastics, and tend to be dose-dependent. Type B reactions result from toxic effects unrelated to the intended pharmacologic actions. The latter effects are often unanticipated (especially with new drugs) and frequently severe and may result from recognized (often immunologic) as well as previously undescribed mechanisms. Type B reactions may occur at low dosages and are often termed dose-independent.

Drugs may increase the frequency of an event that is common in a general population, and this may be especially difficult to recognize; an example is the increase in myocardial infarctions that was seen with the COX-2 inhibitor rofecoxib. Drugs can also cause rare and serious ADRs, such as hematologic abnormalities, arrhythmias, severe skin reactions, or hepatic or renal dysfunction. Prior to regulatory approval and marketing, new drugs are tested in relatively few patients who tend to be less sick and to have fewer concomitant diseases than those patients who subsequently receive the drug therapeutically. Because of the relatively small number of patients studied in clinical trials and the selected nature of these patients, rare ADRs are generally not detected prior to a drug’s approval; indeed, if they are detected, the new drugs are generally not approved. Therefore, physicians need to be cautious in the prescription of new drugs and alert for the appearance of previously unrecognized ADRs.

Elucidating mechanisms underlying ADRs can assist development of safer compounds or allow a patient subset at especially high risk to be excluded from drug exposure. National adverse reaction reporting systems, such as those operated by the FDA (suspected ADRs can be reported online at *http://www.fda.gov/safety/medwatch/default.htm*) and the Committee on Safety of Medicines in Great Britain, can prove useful. The publication or

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

severely hypertensive patients led to development of the drug for hair growth. Sildenafil was initially developed as an antianginal, but its effects to alleviate erectile dysfunction not only led to a new drug indication but also to increased understanding of the role of type 5 phosphodiesterase in

http://www.fda.gov/safety/medwatch/default.htm) and the Committee on Safety of Medicines in Great Britain, can prove useful. The publication or reporting of a newly recognized ADR can in a short time stimulate many similar such reports of reactions that previously had gone unrecognized.

Occasionally, "adverse" effects may be exploited to develop an entirely new indication for a drug. Unwanted hair growth during minoxidil treatment of severely hypertensive patients led to development of the drug for hair growth. Sildenafil was initially developed as an antianginal, but its effects to alleviate erectile dysfunction not only led to a new drug indication but also to increased understanding of the role of type 5 phosphodiesterase in erectile tissue. These examples further reinforce the concept that prescribers must remain vigilant to the possibility that unusual symptoms may reflect unappreciated drug effects.

Some 25–50% of patients make errors in self-administration of prescribed medicines, and these errors can be responsible for ADRs. Similarly, patients commit errors in taking OTC drugs by not reading or following prescribing directions on the containers. Health care providers must recognize that providing directions with prescriptions does not always guarantee compliance.

In hospitals, drugs are administered in a controlled setting, and patient compliance is, in general, ensured. Errors may occur nevertheless—the wrong drug or dose may be given or the drug may be given to the wrong patient—and improved drug distribution and administration systems should help with this problem.

SCOPE OF THE PROBLEM

One estimate in the United Kingdom was that 6.5% of all hospital admissions are due to ADRs and that 2.3% of these patients (0.15%) died as a result. The most common culprit drugs were aspirin, nonsteroidal anti-inflammatory drugs, diuretics, warfarin, ACE inhibitors, antidepressants, opiates, digoxin, steroids, and clopidogrel. One study in the late 1990s suggested that ADRs were responsible for >100,000 in-hospital deaths in the United States, making them the fourth to sixth most common cause of in-hospital death. Another study 10 years later showed no change in this trend.

In hospital, patients receive, on average, 10 different drugs during each hospitalization. The sicker the patient, the more drugs are given, and there is a corresponding increase in the likelihood of ADRs. When <6 different drugs are given to hospitalized patients, the probability of an ADR is ~5%, but if >15 drugs are given, the probability is >40%. Serious ADRs are also well recognized with "herbal" remedies and OTC compounds; examples include kava-associated hepatotoxicity, L-tryptophan-associated eosinophilia-myalgia, and phenylpropanolamine-associated stroke, each of which has caused fatalities.

TOXICITY UNRELATED TO A DRUG'S PRIMARY PHARMACOLOGIC ACTIVITY

Drugs or, more commonly, reactive metabolites generated by CYPs can covalently bind to tissue macromolecules (such as proteins or DNA) to cause tissue toxicity. Because of the reactive nature of these metabolites, covalent binding often occurs close to the site of production, typically the liver.

Acetaminophen

The most common cause of drug-induced hepatotoxicity is acetaminophen overdosage **(Chap. 340)**. Normally, reactive metabolites are detoxified by combining with hepatic glutathione. When glutathione becomes depleted, the metabolites bind instead to hepatic protein, with resultant hepatocyte damage. The hepatic necrosis produced by the ingestion of acetaminophen can be prevented or attenuated by the administration of substances such as *N*-acetylcysteine that reduce the binding of electrophilic metabolites to hepatic proteins. The risk of acetaminophen-related hepatic necrosis is increased in patients receiving drugs such as phenobarbital or phenytoin, which increase the rate of drug metabolism, or ethanol, which exhausts glutathione stores. Such toxicity has even occurred with therapeutic dosages, so patients at risk through these mechanisms should be warned.

Immunologic Reactions

Most pharmacologic agents are haptens, small molecules with low molecular weights (<2000) that are therefore poor immunogens. Generation of an immune response to a drug therefore often requires in vivo activation and covalent linkage to protein, carbohydrate, or nucleic acid.

Drug stimulation of antibody production may mediate tissue injury by several mechanisms. The antibody may attack the drug when the drug is covalently attached to a cell and thereby destroy the cell. This occurs in penicillin-induced hemolytic anemia. Antibody-drug-antigen complexes may be passively adsorbed by a bystander cell, which is then destroyed by activation of complement; this occurs in quinine- and quinidine-induced thrombocytopenia. Heparin-induced thrombocytopenia arises when antibodies against complexes of platelet factor 4 peptide and heparin generate immune complexes that activate platelets; thus, the thrombocytopenia is accompanied by "paradoxical" thrombosis and is treated with thrombin

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Serum sickness **(Chap. 352)** results from the deposition of circulating drug-antibody complexes on endothelial surfaces. Complement activation

thrombocytopenia. Heparin-induced thrombocytopenia arises when antibodies against complexes of platelet factor 4 peptide and heparin generate immune complexes that activate platelets; thus, the thrombocytopenia is accompanied by "paradoxical" thrombosis and is treated with thrombin inhibitors. Drugs or their reactive metabolites may alter a host tissue, rendering it antigenic and eliciting autoantibodies. For example, hydralazine and procainamide (or their reactive metabolites) can chemically alter nuclear material, stimulating the formation of antinuclear antibodies and occasionally causing lupus erythematosus. Drug-induced pure red cell aplasia **(Chap. 102)** is due to an immune-based drug reaction.

Serum sickness **(Chap. 352)** results from the deposition of circulating drug-antibody complexes on endothelial surfaces. Complement activation occurs, chemotactic factors are generated locally, and an inflammatory response develops at the site of complex entrapment. Arthralgias, urticaria, lymphadenopathy, glomerulonephritis, or cerebritis may result. Foreign proteins (vaccines, streptokinase, therapeutic antibodies) and antibiotics are common causes. Many drugs, particularly antimicrobial agents, ACE inhibitors, and aspirin, can elicit anaphylaxis with production of IgE, which binds to mast cell membranes. Contact with a drug antigen initiates a series of biochemical events in the mast cell and results in the release of mediators that can produce the characteristic urticaria, wheezing, flushing, rhinorrhea, and (occasionally) hypotension.

Drugs may also elicit cell-mediated immune responses. One serious reaction is Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN), which can result in death due to T-cell-mediated massive skin sloughing. Another probable immune-mediated drug reaction is the DRESS (drug reaction with eosinophilia and systemic symptoms) syndrome, a rare ADR with a chronic relapsing course, often triggered by antiseizure medications and possibly arising from herpes virus reactivation. As described in **Chap. 68**, specific genetic variants appear necessary but not sufficient to elicit SJS/TEN or DRESS.

While the use of antibodies targeting immune checkpoints is dramatically improving prognosis in many cancers, these agents have also been associated with the unpredictable development of many apparently immune-related ADRs. Some, like colitis or thyroiditis, may be self-limited or medically manageable, while others, notably myocarditis, are rarer but can be rapidly fatal.

DIAGNOSIS AND TREATMENT OF ADVERSE DRUG REACTIONS

The manifestations of drug-induced diseases frequently resemble those of other diseases, and a given set of manifestations may be produced by different and dissimilar drugs. Recognition of the role of a drug or drugs in an illness depends on appreciation of the possible ADRs to drugs in any disease, on identification of the temporal relationship between drug administration and development of the illness, and on familiarity with the common manifestations of the drugs.

A suspected ADR developing after introduction of a new drug naturally implicates that drug; however, it is also important to remember that a drug interaction may be responsible. Thus, for example, a patient on a chronic stable warfarin dose may develop a bleeding complication after introduction of amiodarone; this does not reflect a direct reaction to amiodarone but rather its effect to inhibit warfarin metabolism. Many associations between particular drugs and specific reactions have been described, but there is always a "first time" for a novel association, and any drug should be suspected of causing an ADR if the clinical setting is appropriate.

Illness related to a drug's intended pharmacologic action is often more easily recognized than illness attributable to immune or other mechanisms.

For example, side effects such as cardiac arrhythmias in patients receiving digitalis, hypoglycemia in patients given insulin, or bleeding in patients receiving anticoagulants are more readily related to a specific drug than are symptoms such rash, which may be caused by many drugs or by other factors. Drug fever often escapes initial diagnosis because fever is such a common manifestation of disease.

Electronic listings of ADRs can be useful. However, exhaustive compilations often provide little sense of perspective in terms of frequency and seriousness, which can vary considerably among patients.

Eliciting a drug history from each patient is important for diagnosis. Attention must be directed to OTC drugs and herbal preparations as well as to prescription drugs. Each type can be responsible for ADRs, and adverse interactions may occur between OTC drugs and prescribed drugs. Loss of efficacy of oral contraceptives or cyclosporine with concurrent use of St. John's wort (a P-glycoprotein inducer) is an example (Table 67-2). In addition, it is common for patients to be cared for by several physicians, and duplicative, additive, antagonistic, or synergistic drug combinations may therefore be administered if the physicians are not aware of the patients' drug histories. Electronic health records (EHRs) may help mitigate this problem, but only if all treating physicians use the same EHR system. Medications stopped for inefficacy or adverse effects should be documented to avoid pointless and potentially dangerous reexposure. A frequently overlooked source of additional drug exposure is topical therapy; for example, a patient complaining of bronchospasm may not mention that an ophthalmic beta blocker is being used unless specifically asked. A history of previous ADRs in patients is common. Since these patients have shown a predisposition to drug-induced illnesses, such a history should dictate added caution in

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

agranulocytosis, hemolytic anemia, and thrombocytopenia. For example, both quinine and quinidine can produce platelet agglutination in vitro in the presence of complement and the serum from a patient who has developed thrombocytopenia following use of this drug. Biochemical abnormalities

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

complaining of bronchospasm may not mention that an ophthalmic beta blocker is being used unless specifically asked. A history of previous ADRs in patients is common. Since these patients have shown a predisposition to drug-induced illnesses, such a history should dictate added caution in prescribing new drugs.

Laboratory studies may include demonstration of serum antibody in some persons with drug allergies involving cellular blood elements, as in agranulocytosis, hemolytic anemia, and thrombocytopenia. For example, both quinine and quinidine can produce platelet agglutination in vitro in the presence of complement and the serum from a patient who has developed thrombocytopenia following use of this drug. Biochemical abnormalities such as G6PD deficiency, serum pseudocholinesterase level, or genotyping may also be useful in diagnosis, especially after an ADR has occurred in the patient or a family member **(Chap. 68)**.

Once an ADR is suspected, discontinuation of the suspected drug followed by disappearance of the reaction is presumptive evidence of a drug-induced illness. Confirming evidence may be sought by cautiously reintroducing the drug and seeing if the reaction reappears. However, that should be done only if confirmation would be useful in the future management of the patient. Because rechallenge does carry risks, it is generally avoided unless the suspected culprit drug is critical to the patient's care. When the reaction is thought to be immunologic, challenge is generally avoided. With concentration-dependent ADRs, lowering the dosage may cause the reaction to disappear, and raising it may cause the reaction to reappear. Serious immunologically mediated ADRs have been treated with high-dose steroids; other immunosuppressive agents such as rituximab, infliximab, or mycophenolate mofetil; or plasmapheresis.

If the patient is receiving many drugs when an ADR is suspected, the drugs likeliest to be responsible can usually be identified; this should include both potential culprit agents as well as drugs that alter their elimination. All drugs may be discontinued at once or, if this is not practical, discontinued one at a time, starting with the ones most suspect, and the patient observed for signs of improvement. The time needed for a concentration-dependent ADR to disappear depends on the time required for the concentration to fall below the range associated with the ADR; that, in turn, depends on the initial blood level and on the rate of elimination or metabolism of the drug. Adverse effects of drugs with long half-lives or those not directly related to serum concentration may take a considerable time to disappear.

THE DRUG DEVELOPMENT PROCESS

Drug therapy is an ancient feature of human culture. The first treatments were plant extracts discovered empirically to be effective for indications like fever, pain, or breathlessness. This symptom-based empiric approach to drug development was supplanted in the twentieth century by identification of compounds targeting more fundamental biologic processes, such as bacterial growth or elevated blood pressure. The term "magic bullet," coined by Paul Ehrlich to describe the search for effective compounds for syphilis, captures the essence of the hope that understanding basic biologic processes will lead to highly effective new therapies.

A common starting point for the development of many widely used modern therapies has been basic biologic discovery that implicates potential target molecules: examples of such target molecules include HMG-CoA reductase, a key step in cholesterol biosynthesis, or the *BRAF* V600E mutation that appears to drive the development of some malignant melanomas and other tumors. The development of compounds targeting these molecules has not only revolutionized treatment for diseases such as hypercholesterolemia or malignant melanoma, but has also revealed new biologic features of disease. Thus, for example, initial spectacular successes with vemurafenib (which targets *BRAF* V600E) were followed by near-universal tumor relapse, strongly suggesting that inhibition of this pathway alone would be insufficient for tumor control. This reasoning, in turn, supports a view that many complex diseases will not lend themselves to cure by targeting a single magic bullet, but rather single drugs or combinations that attack multiple pathways whose perturbation results in disease. The use of combination therapy in settings such as hypertension, tuberculosis, HIV infection, and many cancers highlights the potential for such a "systems biology" view of drug therapy.

A common approach in contemporary drug development is to start with a high-throughput screening procedure to identify "lead" chemical(s) modulating the activity of a potential drug target. The next step is application of increasingly sophisticated medicinal chemistry-based modification of the "lead" to develop compounds with specificity for the chosen target, lack of "off-target" effects, and pharmacokinetic properties suitable for human use (e.g., consistent bioavailability, long elimination half-life, and no high-risk pharmacokinetic features). Drug evaluation in human subjects then proceeds from initial safety and tolerance (phase 1) to dose finding (phase 2) and then to large efficacy trials (phase 3). This is a very expensive process, and the vast majority of lead compounds fail at some point. Thus, new approaches to identify likely successes and failures early are needed. One idea, described further in **Chap. 68**, is to use genomic and other high-throughput profiling approaches not only to identify new drug targets but also to identify disease subsets for which drugs approved for other indications might be "repurposed," thereby avoiding the costly development process.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

an individual's response to drug treatment. Molecular pharmacology, pharmacokinetics, genetics, clinical trials, and the educated prescriber all contribute to this process. No drug response should ever be termed *idiosyncratic*; all responses have a mechanism whose understanding will help

SUMMARY

Modern clinical pharmacology aims to replace empiricism in the use of drugs with therapy based on in-depth understanding of factors that determine an individual's response to drug treatment. Molecular pharmacology, pharmacokinetics, genetics, clinical trials, and the educated prescriber all contribute to this process. No drug response should ever be termed *idiosyncratic*; all responses have a mechanism whose understanding will help guide further therapy with that drug or successors. This rapidly expanding understanding of variability in drug actions makes the process of prescribing drugs increasingly daunting for the practitioner. However, fundamental principles should guide this process:

- The benefits of drug therapy, however defined, should always outweigh the risk.
- The smallest dosage necessary to produce the desired effect should be used.
- The number of medications and doses per day should be minimized.
- Although the literature is rapidly expanding, accessing it is becoming easier; electronic tools to search databases of literature and unbiased opinion will become increasingly commonplace.
- Genetics play a role in determining variability in drug response and may become a part of clinical practice.
- EHR and pharmacy systems will increasingly incorporate prescribing advice, such as indicated medications not used; unindicated medications being prescribed; and potential dosing errors, drug interactions, or genetically determined drug responses.
- Prescribers should be particularly wary when adding or stopping specific drugs that are especially liable to provoke interactions and adverse drug reactions.
- Prescribers should use only a limited number of drugs, with which they are thoroughly familiar.

FURTHER READING

Barrett JS et al: Challenges and opportunities in the development of medical therapies for pediatric populations and the role of extrapolation. Clin Pharmacol Ther 103:419, 2018. [PubMed: 29277892]

Holford N: Pharmacodynamic principles and the time course of immediate drug effects. Transl Clin Pharmacol 25:157, 2017. [PubMed: 32095468]

Macrae CA et al: The future of cardiovascular therapeutics. Circulation 133:2610, 2016. [PubMed: 27324356]

Mueller KT et al: The role of clinical pharmacology across novel treatment modalities. Clin Pharmacol Ther 108:413, 2020. [PubMed: 31999830]

Sultana J et al: Clinical and economic burden of adverse drug reactions. J Pharmacol Pharmacother 4:S73, 2013. [PubMed: 24347988]

Zamek-gliszczynski MJ et al: Transporters in drug development: 2018 ITC recommendations for transporters of emerging clinical importance. Clin Pharmacol Ther 104:890, 2018. [PubMed: 30091177]

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 67: Principles of Clinical Pharmacology, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 68: Pharmacogenomics

Dan M. Roden

INTRODUCTION

The previous chapter discussed mechanisms underlying variability in drug action, highlighting pharmacokinetic and pharmacodynamic pathways to beneficial and adverse drug events. Work in the past several decades has defined how genetic variation can play a prominent role in modulating these pathways. Initial studies described unusual drug responses due to single genetic variants in individual subjects, defining the field of pharmacogenetics. A more recent view extends this idea to multiple genetic variants across populations, and the term “pharmacogenomics” is often used. Understanding the role of genetic variation in drug response could improve the use of current drugs, avoid drug use in those at increased risk for adverse drug reactions (ADRs), guide development of new drugs, and even be used as a lens through which to understand mechanisms of diseases themselves. This chapter will outline the principles of pharmacogenomics, the evidence as currently available that genetic factors play a role in variable drug actions, and areas of controversy and ongoing work.

PRINCIPLES OF GENETIC VARIATION AND DRUG RESPONSE

A goal of traditional Mendelian genetics is to identify DNA variants associated with a distinct phenotype in multiple related family members **(See Also Chaps. 466 and 467)**. However, it is unusual for a drug response phenotype to be accurately measured in more than one family member, let alone across a kindred. Some clinical studies have examined drug disposition traits (such as urinary drug excretion after a fixed test dose) in twins and have, in some instances, shown greater concordance in monozygotic compared to dizygotic pairs, supporting a genetic contribution to the trait under study. However, in general, non-family-based approaches are usually used to identify and validate DNA variants contributing to variable drug actions. Both candidate gene and genome-wide studies have been used, and as with any genomic study, results require replication before they should be accepted as valid.

Types of Genetic Variants Influencing Drug Response

The most common type of genetic variant is a single nucleotide polymorphism (SNP), and nonsynonymous SNPs (i.e., those that alter primary amino acid sequence encoded by a gene) are a common cause of variant function in genes regulating drug responses, often termed *pharmacogenes* **(Table 68-1)**. Small insertions and deletions can similarly alter protein function or lead to functionally important splice variation. Examples of synonymous coding region variants altering pharmacogene function have also been described; the postulated mechanism is an alteration in the rate of RNA translation, and hence in folding of the nascent protein. Variation in pharmacogene promoters has been described, and copy number variation (gene deletion or multiple copies of the same gene) is also well described.

TABLE 68-1

Examples of Genetic Variation and Ancestry

STRUCTURAL VARIANT	EXAMPLE		FUNCTIONAL EFFECT	MINOR ALLELE FREQUENCY (%)[a]		
	COMMON NAME	dbSNP		EUROPEAN	AFRICAN	EAST ASIAN
Single nucleotide polymorphism (SNP) (or	CYP2C9*2	rs1799853	R144C: Reduction of function	12.7	2.4	b
	CYP2C19*2	rs4244285	Splicing defect: Loss of function	14.8	18.1	31.0

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

Single nucleotide polymorphism (SNP) (or single nucleotide variant, SNV)	CYP2C9*2	rs1799853	R144C: Reduction of function	12.7	2.4	b
	CYP2C9*3	rs1057910	I359L: Loss of function	6.9	1.3	3.4
	CYP2C9*8	rs7900194	R150H: Reduction of function	b	5.6	b
	CYP2C19*2	rs4244285	Splicing defect: Loss of function	14.8	18.1	31.0
	CYP2C19*3	rs4986893	Premature stop: Loss of function	b	b	6.7
	CYP2C19*17	rs12248560	Gain of function	45	45	<5
	CYP2D6*4c	rs3892097	Splicing defect: Loss of function	23.1	11.9	0.4
	CYP2D6*10c	Multiple SNPs define CYP2D6*10 (reduction of function allele):				
		rs1065852	P34S	24.9	15.1	59.1
		rs1135840	S486T			
	CYP3A5*3	rs776746	Splicing defect: Loss of function	90	33	85
	VKORC1*2	rs9923231	Promoter variant associated with decreased warfarin dose	39	11	91
	VKORC1	rs61742245	D36Y: Reduction of function, associated with increased warfarin dose	5% in East Africa, Middle East, Oceania; rare elsewhere		
	ABCB1	rs1045642	Synonymous variant; may affect mRNA stability and protein folding	47.2	79.8	62.5
Insertion/deletion	UGT1A1*28		Reduction of function promoter variant (7 TA repeats versus 6 repeats in reference allele); homozygotes have Gilbert's syndrome	31.6	39.1	14.8
Multiple variants constituting specific haplotypes	HLA-B*15:01		Predispose to immunologically mediated adverse drug reactions	b	b	5
	HLA-B*57:01			6.8	1.0	1.6
Gene deletion	CYP2D6*5		Loss of function	2.7	6	5.6
Gene duplication	CYP2D6*1xN	Duplication of normal allele	Ultra-rapid metabolizer phenotype	0.8	1.5	0.3
				Up to 3% in North Africa and the Middle East		
	CYP2D6*4xN	Duplication of loss of function allele	Extensive or poor metabolizer phenotype, depending on the opposite allele	0.3	1.4	b

Note: Allele frequencies from https://gnomad.broadinstitute.org/ and https://cpicpgx.org/.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Table 68-1 lists examples of individual types of genomic variation and the impact they can have on function of pharmacogenes. Multiple genotyping

Note: Allele frequencies from *https://gnomad.broadinstitute.org/* and *https://cpicpgx.org/*.

[a]Includes heterozygotes and homozygotes. [b]Allele frequency <0.05%. [c]*CYP2D6* is highly polymorphic, and multiple SNPs may be required to define a specific variant. For example, rs1065852 is present in both *4 and *10 variants. See *https://www.pharmvar.org/*.

Table 68-1 lists examples of individual types of genomic variation and the impact they can have on function of pharmacogenes. Multiple genotyping approaches may be needed to detect important variants; for example, SNP assays may fail to detect large gene duplications, and highly polymorphic regions (such as the major histocompatibility locus on chromosome 6 that includes multiple genes of the human leukocyte antigen [HLA] family) are currently best evaluated by sequencing.

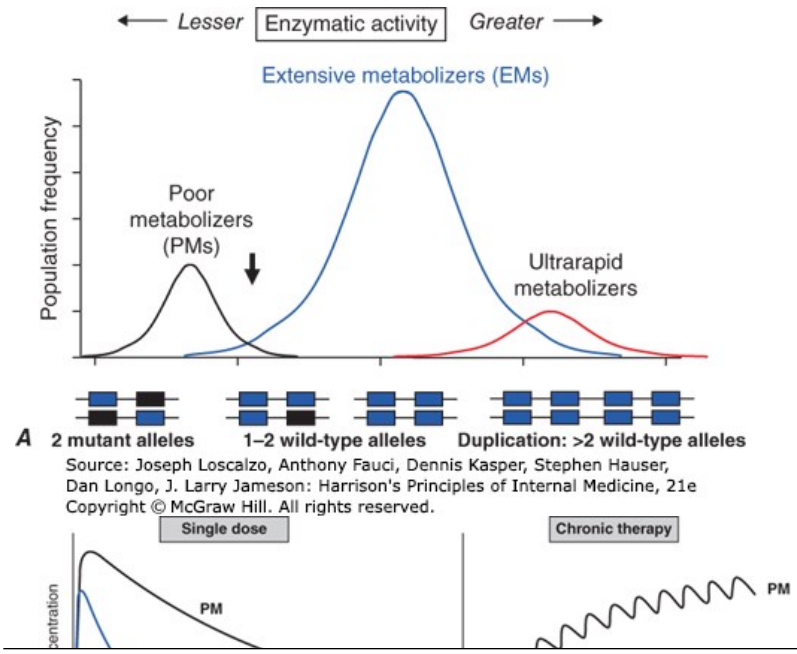
Table 68-1 also highlights the fact that the frequency of important variation across pharmacogenes can vary strikingly by ancestry, with the result that certain ethnic groups may be at unusually high risk of displaying variant response to specific drugs.

Candidate Gene Approaches

Most studies to date have used an understanding of the molecular mechanisms modulating drug action to identify candidate genes in which variants could explain variable drug responses. One very common scenario is that variable drug actions can be attributed to variability in plasma drug concentrations. When plasma drug concentrations vary widely (e.g., more than an order of magnitude), especially if their distribution is non-unimodal as in **Fig. 68-1**, variants in single genes controlling drug concentrations often contribute. In this case, the most obvious candidate genes are those responsible for drug metabolism and elimination. Other candidate genes are those encoding the target molecules with which drugs interact to produce their effects or molecules modulating that response, including those involved in disease pathogenesis.

FIGURE 68-1

A. Distribution of CYP2D6 metabolic activity across a population. The heavy arrow indicates an antimode, separating poor metabolizer subjects (PMs, black), with two loss-of-function CYP2D6 alleles (black), indicated by the intron-exon structures below the chart. Individuals with one or two functional alleles are grouped together as extensive metabolizers (EMs, blue). Also shown are ultra-rapid metabolizers (UMs, red), with 2–12 functional copies of the gene, displaying the greatest enzyme activity. *(Adapted from M-L Dahl et al: J Pharmacol Exp Ther 274:516, 1995.)* ***B.*** These simulations show the predicted effects of CYP2D6 genotype on disposition of a substrate drug. With a single dose (left), there is an inverse "gene-dose" relationship between the number of active alleles and the areas under the time-concentration curves (smallest in UM subjects; highest in PM subjects); this indicates that clearance is greatest in UM subjects. In addition, elimination half-life is longest in PM subjects. The right panel shows that these single-dose differences are exaggerated during chronic therapy: steady-state concentration is much higher in PM subjects (decreased clearance), as is the time required to achieve steady state (longer elimination half-life).

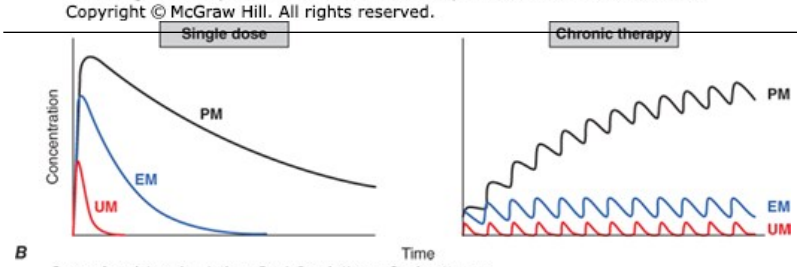


Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Access Provided by:
SILVERCHAIR

Copyright © McGraw Hill. All rights reserved.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Genome-Wide Association Studies

The field has also had some success with "unbiased" approaches such as genome-wide association (GWA) **(Chap. 466)**, particularly in identifying single variants associated with high risk for certain forms of drug toxicity, and in validating the results of candidate gene studies. GWA studies have identified variants in the HLA locus that are associated with high risk for severe skin rashes during treatment with the anticonvulsant carbamazepine and hepatotoxicity with flucloxacillin, an antibiotic never marketed in the United States. A GWA study of simvastatin-associated myopathy identified a single noncoding SNP in *SLCO1B1*, encoding OATP1B1, a drug transporter known to modulate simvastatin uptake into the liver, which accounts for 60% of myopathy risk. African-American subjects are known to have higher dose requirements to achieve stable anticoagulation with warfarin, due in part to variations in *CYP2C9* and *VKORC1*, discussed below. In addition, a GWA study identified novel SNPs near *CYP2C9* that contribute to this effect in African Americans.

GENETIC VARIANTS AFFECTING PHARMACOKINETICS

Clinically important genetic variants have been described in multiple molecular pathways of drug disposition **(Table 68-2)**. A distinct multimodal distribution of drug disposition (as shown in Fig. 68-1) argues for a predominant effect of variants in a single gene in the metabolism of that substrate. Individuals with two alleles (variants) encoding for nonfunctional protein make up one group, often termed *poor metabolizers* (PM phenotype). For most genes, many variants can produce such a loss of function, and assessing whether they are on the same or different alleles (i.e., the *diplotype*) can complicate the use of genotyping in clinical practice. Furthermore, some variants produce only partial loss of function, and the presence of more than one variant may be required to define a specific allele. Individuals with one functional allele, or multiple reduction of function alleles, make up a second group (*intermediate metabolizers*) and may or may not be distinguishable from those with two functional alleles (normal metabolizers, sometimes termed *extensive metabolizers*, EMs). *Ultra-rapid metabolizers* (UMs) with especially high enzymatic activity (occasionally due to gene duplication; Table 68-1 and Fig. 68-1) have also been described for some traits. Many drugs in widespread use can inhibit specific drug disposition pathways **(see Chap. 67, Table 67-1)**, and so EM individuals receiving such inhibitors can respond like PM patients (*phenocopying*). Polymorphisms in genes encoding drug uptake or drug efflux transporters may be other contributors to variability in drug delivery to target sites and, hence, in drug effects.

TABLE 68-2

Genetic Variants and Drug Responses

GENE	DRUGS	EFFECT OF GENETIC VARIANTS[a]
Variants in Drug Metabolism Pathways		
CYP2C9	Losartan	Decreased bioactivation and effects (PMs)
	Warfarin	Decreased dose requirements; possible increased bleeding risk (PMs)
	Phenytoin	Decreased dose requirement (PMs)
CYP2C19	Omeprazole, voriconazole	Decreased effect in EMs
		Consider alternate drug in PMs and alternate drug or dose increase in UMs
		Possible increased bleeding risk in carriers of gain-of-function variants

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION / SYSTEMS

	Celecoxib	Exaggerated effect in PMs
	Clopidogrel	Decreased effect in PMs and IMs Consider alternate drug in PMs and alternate drug or dose increase in IMs Possible increased bleeding risk in carriers of gain-of-function variants
	Citalopram, escitalopram	Choose alternate drug in UMs; reduce dose in PMs
CYP2D6	Codeine, tamoxifen	Decreased bioactivation and drug effects in PMs
	Codeine	Respiratory depression in UMs
	Tricyclic antidepressants[b]	Increased adverse effects in PMs: Consider dose decrease Decreased therapeutic effects in UMs: Consider alternate drug
	Metoprolol, carvedilol, timolol, propafenone	Increased beta blockade in PMs
	Fluvoxamine	Reduce dose or chose alternate drug in PMs
CYP3A5	Tacrolimus, vincristine	Decreased drug concentrations and effect (CYP3A5*3 carriers)
Dihydropyrimidine dehydrogenase (*DPYD*)	Capecitabine, 5-fluorouracil, tegafur	Possible severe toxicity (PMs)
NAT2	Rifampin, isoniazid, pyrazinamide, hydralazine, procainamide	Increased risk of toxicity in PMs
Thiopurine *S*-methyltransferase (*TPMT*)	Azathioprine, 6-mercaptopurine, thioguanine	PMs: Increased risk of bone marrow aplasia EMs: Possible decreased drug action at usual dosages
Uridine diphosphate glucuronosyltransferase (*UGT1A1*)	Irinotecan	PM homozygotes: Increased risk of severe adverse effects (diarrhea, bone marrow aplasia)
	Atazanavir	High risk of hyperbilirubinemia during treatment; can result in drug discontinuation
Pseudocholinesterase (*BCHE*)	Succinylcholine and other muscle relaxants	Prolonged paralysis (autosomal recessive); diagnosis established by genotyping or by measuring serum cholinesterase activity
Variants in Other Genes		
Glucose 6-phosphate dehydrogenase (*G6PD*)	Rasburicase, primaquine, chloroquine	Increased risk of hemolytic anemia in G6PD-deficient subjects
HLA-B*15:02	Carbamazepine	Carriers (1 or 2 alleles) at increased risk of SJS/TEN (mainly Asian subjects)
HLA-B*31:01	Carbamazepine	Carriers (1 or 2 alleles) at increased risk of SJS/TEN and milder skin toxicities (Caucasian and Asian subjects)
HLA-B*15:02	Phenytoin	Carriers (1 or 2 alleles) at increased risk of SJS/TEN

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

IFNL3 (IL28B)	Interferon	Variable response in hepatitis C therapy

ACCESS Medicine®
Access Provided by: SILVERCHAIR INFORMATION SYSTEMS

HLA-B*57:01	Abacavir	Carriers (1 or 2 alleles) at increased risk of SJS/TEN
HLA-B*58:01	Allopurinol	Carriers (1 or 2 alleles) at increased risk of SJS/TEN
IFNL3 (IL28B)	Interferon	Variable response in hepatitis C therapy
SLCO1B1	Simvastatin	Encodes a drug uptake transporter; variant nonsynonymous single nucleotide polymorphism increases myopathy risk especially at higher dosages
VKORC1	Warfarin	Decreased dose requirements with variant promoter haplotype Increased dose requirement in individuals with nonsynonymous loss-of-function variants
ITPA	Ribavirin	Variants modulate risk for hemolytic anemia
RYR1	General anesthetics	Variants predispose to malignant hyperthermia
CFTR	Ivacaftor, lumacaftor	Targeted therapies for cystic fibrosis indicated only in certain genotypes
Variants in Other Genomes (Infectious Agents, Tumors)		
Chemokine C-C motif receptor (CCR5)	Maraviroc	Drug effective only in HIV strains with CCR5 detectible
C-KIT	Imatinib	In gastrointestinal stromal tumors, drug indicated only with c-kit–positive cases
ALK (anaplastic lymphoma kinase)	Crizotinib	Indicated in patients with non-small cell lung cancer and ALK mutations
Her2/neu overexpression	Trastuzumab, lapatinib	Drugs indicated only with tumor overexpression
K-ras mutation	Panitumumab, cetuximab	Lack of efficacy with *KRAS* mutation
Philadelphia chromosome	Dasatinib, nilotinib, imatinib	Decreased efficacy in Philadelphia chromosome–negative chronic myelogenous leukemia

[a]Drug effect in homozygotes unless otherwise specified. [b]Many tricyclic antidepressants and selective serotonin uptake inhibitors are metabolized by CYP2D6, CYP2C19, or both, and some metabolites have pharmacologic activity. See *https://www.pharmgkb.org/view/dosing-guidelines.do*.

Abbreviations: EM, extensive metabolizer (normal enzymatic activity); IM, intermediate metabolizer (heterozygote for loss-of-function allele); PM, poor metabolizer (homozygote for reduced or loss-of-function allele); SJS/TEN, Stevens-Johnson syndrome/toxic epidermal necrolysis; UM, ultra-rapid metabolizer (enzymatic activity much greater than normal, e.g., with gene duplication, Fig. 68-1).

Further data at:

U.S. Food and Drug Administration: *http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm*

Pharmacogenetics Research Network/Knowledge Base: *http://www.pharmgkb.org*

The Clinical Pharmacogenomics Implementation Consortium*: https://www.pharmgkb.org/page/cpic*

Dutch Pharmacogenetics Working Group: *https://www.knmp.nl/patientenzorg/medicatiebewaking/farmacogenetica/pharmacogenetics-1/pharmacogenetics*

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

order of magnitude) among individuals, but nonsynonymous coding region polymorphisms (those that change the encoded amino acid) are rare.

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

CYP3A

Members of the CYP3A family (*CYP3A4*, *CYP3A5*) metabolize the greatest number of drugs in therapeutic use. CYP3A4 activity is highly variable (up to an order of magnitude) among individuals, but nonsynonymous coding region polymorphisms (those that change the encoded amino acid) are rare. Thus, the underlying mechanism likely reflects genetic variation in regulatory regions.

Most subjects of European or Asian origin carry a polymorphism that disrupts splicing in the closely related *CYP3A5* gene. As a result, these individuals display reduced CYP3A5 activity, whereas CYP3A5 activity tends to be greater in subjects of African origin. Decreased efficacy of the antirejection agent tacrolimus in subjects of African origin has been attributed to more rapid CYP3A5-mediated elimination, and a lower risk of vincristine-associated neuropathy has been reported in CYP3A5 "expressers."

CYP2D6

CYP2D6 is second to CYP3A4 in the number of commonly used drugs that it metabolizes. CYP2D6 activity is polymorphically distributed, and 5–10% of European- and African-derived populations (but few Asians) display the PM phenotype (Fig. 68-1). Dozens of loss-of-function variants in *CYP2D6* have been described; the PM phenotype arises in individuals with two such alleles. In addition, UMs with multiple functional copies of *CYP2D6* have been identified especially in East Africa, the Middle East, and Oceania. PMs have slower elimination rates and lower clearance of substrate drugs; as a consequence (Fig. 68-1B), steady-state concentrations are higher and the time taken to achieve steady state is longer than in EMs **(Chap. 67)**. Conversely, UMs display very low steady-state parent drug concentrations and an abbreviated time to steady state.

Codeine is biotransformed by CYP2D6 to the potent active metabolite morphine, so its effects are blunted in PMs and exaggerated in UMs. Deaths due to respiratory depression in children given codeine after tonsillectomy have been attributed to the UM trait, and the U.S. Food and Drug Administration (FDA) has revised the package insert to include a prominent "black box" warning against its use in this setting, and, in fact, forbidding its use in children less than 12 years old. In the case of drugs with beta-blocking properties metabolized by CYP2D6, greater signs of beta blockade (e.g., bronchospasm, bradycardia) have been reported in PM subjects than in EMs. This can be seen not only with orally administered beta blockers such as metoprolol and carvedilol, but also with ophthalmic timolol and with the sodium channel–blocking antiarrhythmic propafenone, a CYP2D6 substrate with beta-blocking properties. UMs may require very high dosages of nortriptyline and other tricyclic antidepressants to achieve a therapeutic effect. Tamoxifen undergoes CYP2D6-mediated biotransformation to an active metabolite, so its efficacy may be in part related to this polymorphism. In addition, the widespread use of selective serotonin reuptake inhibitors (SSRIs) to treat tamoxifen-related hot flashes may also alter the drug's effects because many SSRIs, notably fluoxetine and paroxetine, are also CYP2D6 inhibitors (Table 67-2).

CYP2C19

The PM phenotype for CYP2C19 is common (20%) among Asians and rarer (2–3%) in other populations; the frequency of the PM trait is especially high (>50%) in Oceania. The impact of polymorphic CYP2C19-mediated metabolism has been demonstrated with the proton pump inhibitor omeprazole, where ulcer cure rates with "standard" dosages were much lower in EM patients (29%) than in PMs (100%). Thus, understanding the importance of this polymorphism would have been important in developing the drug, and knowing a patient's *CYP2C19* genotype could improve therapy. CYP2C19 is responsible for bioactivation of the antiplatelet drug clopidogrel, and several large retrospective, and more recently prospective, studies have documented decreased efficacy (e.g., increased myocardial infarction after placement of coronary stents or increased stroke or transient ischemic attacks) among subjects with one or two reductions of function alleles. In addition, some studies suggest that omeprazole and possibly other proton pump inhibitors phenocopy this effect by inhibiting CYP2C19.

CYP2C9

There are common variants in *CYP2C9* that encode proteins with reduction or loss of catalytic function. These variant alleles are associated with increased rates of neurologic complications with phenytoin, hypoglycemia with glipizide, and reduced warfarin dose required to maintain stable anticoagulation. Rare patients homozygous for loss-of-function alleles may require very low warfarin dosages. Up to 50% of the variability in steady-state warfarin dose requirement is attributable to polymorphisms in *CYP2C9* and in the promoter of *VKORC1*, which encodes the warfarin target with lesser contributions by genes such as *CYP4F2* controlling vitamin K metabolism. The angiotensin receptor blocker losartan is a prodrug that is bioactivated by CYP2C9; as a result, PMs and those receiving inhibitor drugs may display little response to therapy.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

when exposed to the substrate anticancer drug 5-fluorouracil (5-FU), as well as to capecitabine and tegafur, which are metabolized to 5-FU. Dose reductions have been recommended in intermediate metabolizers.

DPYD

Individuals homozygous for loss-of-function alleles in dihydropyrimidine dehydrogenase, encoded by *DPYD*, are at increased risk for severe toxicity when exposed to the substrate anticancer drug 5-fluorouracil (5-FU), as well as to capecitabine and tegafur, which are metabolized to 5-FU. Dose reductions have been recommended in intermediate metabolizers.

Transferase Variants

Thiopurine *S*-methyltransferase (TPMT) bioinactivates the antileukemic drug 6-mercaptopurine (6-MP), and 6-MP is itself an active metabolite of the immunosuppressive azathioprine. Homozygotes for alleles encoding inactive TPMT (1/300 individuals) predictably exhibit severe and potentially fatal pancytopenia on standard doses of azathioprine or 6-MP. On the other hand, homozygotes for fully functional alleles may display less anti-inflammatory or antileukemic effect with standard doses of the drugs. GWA studies have also identified loss-of-function variants in *NUDT15* that reduce degradation of thiopurine metabolites and, thereby, also increase risk of excessive myelosuppression.

N-acetylation is accomplished by hepatic *N*-acetyl transferase (NAT), which represents the activity of two genes, *NAT1* and *NAT2*. Both enzymes transfer an acetyl group from acetyl coenzyme A to the drug; polymorphisms in *NAT2* are thought to underlie individual differences in the rate at which drugs are acetylated and thus define “rapid acetylators” and “slow acetylators.” Slow acetylators make up ~50% of European and African populations but are less common among East Asians. Slow acetylators have an increased incidence of the drug-induced lupus syndrome during procainamide and hydralazine therapy and of hepatitis with isoniazid.

Individuals homozygous for a common promoter polymorphism that reduces transcription of uridine diphosphate glucuronosyltransferase (*UGT1A1*) have benign hyperbilirubinemia (Gilbert’s syndrome; **Chap. 337**). This variant has also been associated with diarrhea and increased bone marrow depression with the antineoplastic prodrug irinotecan, whose active metabolite is normally detoxified by UGT1A1-mediated glucuronidation. The antiretroviral atazanavir is a UGT1A1 inhibitor, and individuals with the Gilbert’s variant develop higher bilirubin levels during treatment. While this is benign, the hyperbilirubinemia can complicate clinical care because it may raise the question of whether coexistent hepatic injury is present.

Transporter Variants

The risk for myotoxicity with simvastatin and possibly other statins appears increased with variants in *SLCO1B1*. Variants in *ABCB1*, encoding the drug efflux transporter P-glycoprotein, may increase digoxin toxicity. Variants in the uptake transporters *MATE1* and *MATE2* have been reported to modulate metformin’s glucose-lowering activity.

GENETIC VARIANTS AFFECTING PHARMACODYNAMICS

A variant in the *VKORC1* promoter, especially common in Asian subjects (Table 68-1), reduces transcriptional activity and warfarin dose requirement. Multiple polymorphisms identified in the β_2-adrenergic receptor appear to be linked to specific drug responses in asthma and congestive heart failure, diseases in which β_2-receptor function might be expected to determine drug response. Polymorphisms in the β_2-receptor gene have also been associated with response to inhaled β_2-receptor agonists, while those in the β_1-adrenergic receptor gene have been associated with variability in heart rate slowing and blood pressure lowering. In addition, in heart failure, the arginine allele of the common β_1-adrenergic receptor gene polymorphism R389G has been associated with decreased mortality and decreased incidence of atrial fibrillation during treatment with the investigational beta blocker bucindolol.

Drugs may also interact with genetic pathways of disease to elicit or exacerbate symptoms of the underlying conditions. In the porphyrias, CYP inducers are thought to increase the activity of enzymes proximal to the deficient enzyme, exacerbating or triggering attacks **(Chap. 416)**. Deficiency of glucose-6-phosphate dehydrogenase (G6PD), most often in individuals of African, Mediterranean, or South Asian descent, increases the risk of hemolytic anemia in response to the antimalarial primaquine **(Chap. 100)** and the uric acid-lowering agent rasburicase, which does not cause hemolysis in patients with normal amounts of the enzyme. Patients with mutations in *RYR1* encoding the skeletal muscle intracellular release calcium (also termed type 1 ryanodine receptor) are asymptomatic until exposed to certain general anesthetics, which can trigger the rare syndrome of malignant hyperthermia. Certain antiarrhythmics and other drugs can produce marked QT prolongation and torsades de pointes **(Chap. 246)**, and in a minority of affected patients, this adverse effect represents unmasking of previously subclinical congenital long QT syndrome.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

be linked to specific HLA alleles (Table 68-2). Cases of drug-induced hepatotoxicity and of the drug rash with eosinophilia and systemic symptoms (DRESS) syndrome have also been linked to variants in this region. The frequency of risk alleles often varies by ancestry (Table 68-1). The HLA risk

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

a minority of affected patients, this adverse effect represents unmasking of previously subclinical congenital long QT syndrome.

Immunologically Mediated Drug Reactions

The Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN) is a potentially fatal skin and systemic reaction now increasingly recognized to be linked to specific HLA alleles (Table 68-2). Cases of drug-induced hepatotoxicity and of the drug rash with eosinophilia and systemic symptoms (DRESS) syndrome have also been linked to variants in this region. The frequency of risk alleles often varies by ancestry (Table 68-1). The HLA risk alleles appear to be necessary but not sufficient to elicit these reactions. For example, HLA-B*57:01 is a risk allele for abacavir-related SJS/TEN and flucloxacillin-related hepatotoxicity. However, while 55% of abacavir-exposed subjects will develop a reaction, only 1/10,000 subjects exposed to flucloxacillin develop hepatotoxicity. Thus, a third factor, the nature of which has not yet been established, seems necessary.

Tumor and Infectious Agent Genomes

The actions of drugs used to treat infectious or neoplastic disease may be modulated by variants in these nonhuman germline genomes. Genotyping tumors is a rapidly evolving approach to target therapies to underlying mechanisms and to avoid potentially toxic therapy in patients who would derive no benefit **(Chap. 71)**. Trastuzumab, which potentiates anthracycline-related cardiotoxicity, is ineffective in breast cancers that do not express the Herceptin receptor. Imatinib targets a specific tyrosine kinase, BCR-Abl1, that is generated by the translocation that creates the Philadelphia chromosome typical of chronic myelogenous leukemia (CML). Imatinib is also an inhibitor of another kinase, c-kit, and the drug is remarkably effective in c-kit-driven cancer, such as gastrointestinal stromal tumors **(Chap. 71)**. Vemurafenib does not inhibit wild-type *BRAF* but is active against the V600E mutant form of the kinase. Crizotinib is highly effective in non-small cell lung cancers harboring anaplastic lymphoma kinase (ALK) mutations.

INCORPORATING PHARMACOGENETIC INFORMATION INTO CLINICAL PRACTICE

The discovery of common variant alleles with relatively large effects on drug response raises the prospect that these variants could be used to guide therapy. Desired outcomes could be better ways of choosing likely effective drugs and dosages, or avoiding drugs that are likely to produce severe adverse drug events or be ineffective in individual subjects. Indeed, the FDA now incorporates pharmacogenetic data into package inserts meant to guide prescribing. A decision to adopt pharmacogenetically guided dosing for a given drug depends on multiple factors. The most important are the magnitude and clinical importance of the genetic effect and the strength of evidence linking genetic variation to variable drug effects (e.g., anecdote versus post-hoc analysis of clinical trial data versus randomized clinical trial [RCT]). The evidence can be strengthened if statistical arguments from clinical trial data are complemented by an understanding of underlying physiologic mechanisms. Cost versus expected benefit may also be a factor.

Point of Care Versus Preemptive Approaches

Two approaches to pharmacogenetic implementation have been put in place at “early adopter” institutions and are currently being evaluated. In the first, variant-specific assays are ordered at the time of drug prescription and delivered rapidly (often within an hour or two), and the results are then used to guide therapy with that specific drug. The alternative to this “point-of-care” approach is a “preemptive” approach in which pharmacogenetic testing for large numbers of potential variants across many drugs is undertaken prior to prescription of any such drug. The data are then available in electronic health record (EHR) systems and coupled to real-time clinical decision support (CDS). When a drug whose effects are known to be influenced by pharmacogenetic variants is prescribed, the EHR system looks up whether variants likely to affect response are present; if so, CDS will alert health care providers that an alternate drug or a different dose may be required.

Challenges

There are multiple challenges in putting in place either system. Assay validity and reproducibility have been issues in the past, but are less likely now. National consortia are now being put in place to develop standards for pharmacogenetic CDS. While common variants in genes such as those listed in Table 68-1 have been clearly associated with variable drug responses, the effect of rare variants, now readily discoverable by large-scale sequencing, is unknown. The extent to which a dose adjustment might be recommended may vary depending on whether zero, one, or two variant alleles are present, and whether such variants are reduction of function, loss of function, or gain of function. The Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Dutch Pharmacogenetics Working Group have developed and published guidelines for multiple drug-gene pairs focusing on the question of what might be an appropriate drug dose adjustment given the availability of genetic data. These resources do not directly address the question of when or how such genetic testing should be undertaken.

Developing Evidence That Pharmacogenetic Testing Alters Drug Outcomes

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

genetic testing as a guide to prescribing; HLA-B*57:01 testing for abacavir is an example described below. In other situations, the arguments are less compelling: the magnitude of the genetic effect may be smaller, the consequences may be less serious, alternate therapies may be available, or the

Developing Evidence That Pharmacogenetic Testing Alters Drug Outcomes

A major issue is whether pharmacogenetic testing affects important drug response outcomes. When the evidence is compelling, alternate therapies are not available, and there are clear recommendations for dosage adjustment in subjects with variants, there is a strong argument for deploying genetic testing as a guide to prescribing; HLA-B*57:01 testing for abacavir is an example described below. In other situations, the arguments are less compelling: the magnitude of the genetic effect may be smaller, the consequences may be less serious, alternate therapies may be available, or the drug effect may be amenable to monitoring by other approaches.

One school argues that the physiology and pharmacology are known and that RCTs are, therefore, unnecessary (and conceivably unethical). The analogy is sometimes drawn to well-recognized dose adjustment of renally excreted drugs in the presence of renal dysfunction. RCTs have not been conducted and the idea of such dose adjustment is well accepted in the medical community and recommended in FDA-approved drug labels. Others have argued that the effect of genetic variants is generally modest and variability in drug actions has many nongenetic sources, so genetic testing might provide marginal benefit at best.

Efforts to demonstrate the value of pharmacogenetic testing have met with mixed results. An RCT clearly showed that HLA-B*57:01 testing eliminates SJS/TEN due to abacavir. Similarly, regulatory authorities in some countries in Southeast Asia mandated HLA-B*15:02 testing prior to initiation of carbamazepine; however, in this case, an unfortunate outcome in some jurisdictions was that prescribers stopped using carbamazepine, often substituting phenytoin (another drug associated with SJS/TEN), so the incidence of the severe ADR was unchanged.

RCTs evaluating the effect of using pharmacogenetically guided therapy to optimize warfarin treatment have shown either no effect or a modest benefit of incorporating genetic information into prescribing the drug. Initial RCTs focused on time in therapeutic range in the first 4–12 weeks of treatment, whereas one more recent trial demonstrated that genotype-guided therapy could reduce the frequency of over-anticoagulation. Retrospective analyses of bleeding cases versus nonbleeding controls in EHRs and administrative databases have suggested a role for CYP2C9*3 or for the V433M variant in *CYP4F2* in mediating this risk.

Two large trials have randomized patients with acute coronary syndromes to newer antiplatelet therapies (ticagrelor or prasugrel) or clopidogrel if *CYP2C19* variants were absent; in one, clopidogrel was superior, and in the second, a trend in the same direction, which did not reach the prespecified endpoint, was observed.

New effective alternate therapies to warfarin and clopidogrel that appear to lack important pharmacogenetic variants have emerged. One approach to therapy, therefore, is to use pharmacogenetic testing to identify subjects in whom variants are absent and therefore standard doses of the conventional inexpensive drugs are likely to be effective and to reserve alternate more expensive therapies for subjects likely to have variant responses to warfarin or clopidogrel.

GENETICS AND DRUG DEVELOPMENT

Genetic tools are now being increasingly used to identify or validate new drug targets. Initial studies suggest that a new drug development program is more likely to succeed if evidence from human genetics supports the role of a possible drug target in disease pathogenesis and suggests that the risk of toxicity due to high-risk pharmacokinetics or other mechanisms is small. Furthermore, studies of the relationships between variants in genes encoding drug target molecules and a range of phenotypes (e.g., those in EHRs) are being used for drug “repurposing,” identifying new indications for existing drugs.

Finding Protective Alleles Can Identify Drug Targets

One example of using genetics to identify a new drug target started with the discovery that very rare gain-of-function variants in *PCSK9* are a rare cause of familial hypercholesterolemia. Subsequently, population studies showed that carriers of loss-of-function SNPs (2.5% of African Americans) had decreased low-density lipoprotein cholesterol, decreased incidence of coronary artery disease, and no deleterious consequences in other organ systems. These data triggered the development of PCSK9 monoclonal antibodies, which were marketed <10 years after the initial population studies. Other targets implicated by similar population genetic studies include HSD17B13 for prevention of chronic liver disease, SLC30A8 for the prevention of type 2 diabetes, and APOC3 for hypertriglyceridemia. Discovering rare protective alleles may require very large data sets (>100,000), such as EHR systems coupled to DNA biobanks or epidemiologic cohorts like the UK Biobank.

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

rapid emergence of drug resistance, often after extraordinary initial responses. For example, 40% of melanomas appear to be driven by the V600E mutant form of *BRAF*, and the specific inhibitor vemurafenib can produce clinically spectacular remission. However, durable responses are rare, and it

systems coupled to DNA biobanks or epidemiologic cohorts like the UK Biobank.

Cancer

In cancer, tumor sequencing has identified new targets for drug development, often constitutively active kinases. A problem in this area has been the rapid emergence of drug resistance, often after extraordinary initial responses. For example, 40% of melanomas appear to be driven by the V600E mutant form of *BRAF*, and the specific inhibitor vemurafenib can produce clinically spectacular remission. However, durable responses are rare, and it is now apparent that combination therapy, often with inhibitors of the MEK pathway, can provide improved therapy. Another approach that is rapidly gaining wide use in cancer involves drugs that reverse immune system inhibition **(Chap. 73)**. In some patients, the release of this "brake" can provide durable remissions, whereas in others, severe adverse events, including colitis, pneumonitis, and myocarditis, have been reported. Understanding the mechanisms underlying variability to these therapies is a major emerging challenge in the field.

Using Multiple Data Types

The development of methods to understand associations across multiple large data sets is another approach that is being explored in drug development. For example, a GWA study of risk of rheumatoid arthritis identified multiple risk loci, and many encode proteins that are known targets for intervention in the disease. Interestingly, others encode proteins that are targets for drugs used in other conditions, such as certain cancers, raising the question of whether such drugs could be "repurposed" for rheumatoid arthritis.

While the field has, to date, focused on individual high effect size variants (that are often common in a population), newer approaches combining many (dozens to millions) common variants into polygenic risk scores to predict drug responses are also being explored. An extension of this approach is the broader issue of systems pharmacology, in which multiple sources of data are used to identify potential molecules or pathways that would be amenable to treatment, by new drugs or by existing agents, using analysis of genomic, transcriptomic, proteomic, and other large data sets. Similar approaches are being developed to predict toxicity expected from targeting specific genes or disease pathways.

SUMMARY

The science of pharmacogenomics has evolved from isolated examples of rare adverse drug actions to a more comprehensive view of the role of genetic variation in mediating the effects of most drugs. Current principles include:

- Genetic variants with an important effect on drug actions can be common, and their frequencies often vary by ancestry.
- One common mechanism is modulation of drug concentrations.
- No practitioner can be expected to remember all variants important for all drugs. Electronic data systems can now be accessed to describe this information. Ultimately, this information will be used by linking individual pharmacogenetic data to smart EHR systems.
- Incorporating genetic approaches into drug development projects holds the promise of more rapid development of targeted, safe, and effective therapies.

FURTHER READING

Chenoweth MJ et al: Global pharmacogenomics within precision medicine: Challenges and opportunities. Clin Pharmacol Ther 107:57, 2020. [PubMed: 31696505]

Diogo D et al: Phenome-wide association studies across large population cohorts support drug target validation. Nat Commun 9:4285, 2018. [PubMed: 30327483]

Luzum JA et al: The Pharmacogenomics Research Network Translational Pharmacogenetics Program: Outcomes and metrics of pharmacogenetic implementations across diverse healthcare systems. Clin Pharmacol Ther 102:502, 2017. [PubMed: 28090649]

Osanlou O et al: Pharmacogenetics of adverse drug reactions. Adv Pharmacol 83:155, 2018. [PubMed: 29801574]

Relling MV et al: The clinical pharmacogenetics implementation consortium: 10 years later. Clin Pharmacol Ther 107:171, 2020. [PubMed: 31562822]

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Relling MV et al: The clinical pharmacogenetics implementation consortium: 10 years later. Clin Pharmacol Ther 107:171, 2020. [PubMed: 31562822]

Roden DM et al: Pharmacogenomics. Lancet 394:521, 2019. [PubMed: 31395440]

Downloaded 2022-1-3 5:51 P Your IP is 141.216.78.40
Chapter 68: Pharmacogenomics, Dan M. Roden
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 77: Head and Neck Cancer

Everett E. Vokes

INTRODUCTION

Epithelial carcinomas of the head and neck arise from the mucosal surfaces in the head and neck and typically are squamous cell in origin. This category includes tumors of the paranasal sinuses, the oral cavity, and the nasopharynx, oropharynx, hypopharynx, and larynx. Tumors of the salivary glands differ from the more common carcinomas of the head and neck in etiology, histopathology, clinical presentation, and therapy. They are rare and histologically highly heterogeneous. **Thyroid malignancies are described in Chap. 385.**

INCIDENCE AND EPIDEMIOLOGY

The number of new cases of head and neck cancers (oral cavity, pharynx, and larynx) in the United States was estimated at 65,630 in 2020, accounting for about 4% of adult malignancies; estimated deaths were 14,500. The worldwide incidence exceeds half a million cases annually. In North America and Europe, the tumors usually arise from the oral cavity, oropharynx, or larynx. The incidence of oropharyngeal cancers is increasing in recent years, especially in Western countries. Nasopharyngeal cancer is more commonly seen in the Mediterranean countries and in the Far East, where it is endemic in some areas.

ETIOLOGY AND GENETICS

Alcohol and tobacco use are the most significant environmental risk factors for head and neck cancer, and when used together, they act synergistically. Smokeless tobacco is an etiologic agent for oral cancers. Other potential carcinogens include marijuana and occupational exposures such as nickel refining, exposure to textile fibers, and woodworking.

Some head and neck cancers have a viral etiology. Epstein-Barr virus (EBV) infection is frequently associated with nasopharyngeal cancer, especially in endemic areas of the Mediterranean and Far East. EBV antibody titers can be measured to screen high-risk populations and are under investigation to monitor treatment response. Nasopharyngeal cancer has also been associated with consumption of salted fish and indoor pollution.

In Western countries, the human papillomavirus (HPV) is associated with a rising incidence of tumors arising from the oropharynx, that is, the tonsillar bed and base of tongue. Over 50% of oropharyngeal tumors are caused by HPV in the United States, and in many urban centers, this proportion is higher. HPV 16 is the dominant viral subtype, although HPV 18 and other oncogenic subtypes are seen as well. Alcohol- and tobacco-related cancers, on the other hand, have decreased in incidence. HPV-related oropharyngeal cancer frequently occurs in a younger patient population and is associated with increased numbers of sexual partners and oral sexual practices. It is associated with a better prognosis, especially for nonsmokers. Vaccination with the nine-valent HPV vaccine may prevent the disease in high-risk populations.

Dietary factors may contribute. The incidence of head and neck cancer is higher in people with the lowest consumption of fruits and vegetables. Certain vitamins, including carotenoids, may be protective if included in a balanced diet. Supplements of retinoids, such as *cis*-retinoic acid, have not been shown to prevent head and neck cancers (or lung cancer) and may increase the risk in active smokers. No specific risk factors or environmental carcinogens have been identified for salivary gland tumors.

HISTOPATHOLOGY, CARCINOGENESIS, AND MOLECULAR BIOLOGY

Squamous cell head and neck cancers are divided into well-differentiated, moderately well-differentiated, and poorly differentiated categories. Poorly differentiated tumors have a worse prognosis than well-differentiated tumors. For nasopharyngeal cancers, the less common differentiated squamous cell carcinoma is distinguished from nonkeratinizing and undifferentiated carcinoma (lymphoepithelioma) that contains infiltrating lymphocytes and

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 77: Head and Neck Cancer, Everett E. Vokes
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

aerodigestive tract). Most parotid tumors are benign, but half of submandibular and sublingual gland tumors and most minor salivary gland tumors are malignant. Malignant tumors include mucoepidermoid and adenoid cystic carcinomas and adenocarcinomas.

differentiated tumors have a worse prognosis than well-differentiated tumors. For nasopharyngeal cancers, the less common differentiated squamous cell carcinoma is distinguished from nonkeratinizing and undifferentiated carcinoma (lymphoepithelioma) that contains infiltrating lymphocytes and is commonly associated with EBV.

Salivary gland tumors can arise from the major (parotid, submandibular, sublingual) or minor salivary glands (located in the submucosa of the upper aerodigestive tract). Most parotid tumors are benign, but half of submandibular and sublingual gland tumors and most minor salivary gland tumors are malignant. Malignant tumors include mucoepidermoid and adenoid cystic carcinomas and adenocarcinomas.

The mucosal surface of the entire pharynx is exposed to alcohol- and tobacco-related carcinogens and is at risk for the development of a premalignant or malignant lesion. Erythroplakia (a red patch) or leukoplakia (a white patch) can be histopathologically classified as hyperplasia, dysplasia, carcinoma in situ, or carcinoma. However, most head and neck cancer patients do not present with a known history of premalignant lesions. Multiple synchronous or metachronous cancers can also be observed. In fact, over time, patients with treated early-stage tobacco- and alcohol-related head and neck cancer are at greater risk of dying from a second malignancy than from a recurrence of the primary disease.

Second head and neck malignancies are usually not therapy induced; they reflect the exposure of the upper aerodigestive mucosa to the same carcinogens that caused the first cancer. These second primaries develop in the head and neck area, the lung, or the esophagus. Thus, computed tomography (CT) screening for lung cancer in heavy smokers who have already developed a head and neck cancer is recommended. Rarely, patients can develop a radiation therapy–induced sarcoma after having undergone prior radiotherapy for a head and neck cancer.

Much progress has been made in describing the molecular features of head and neck cancer. These features have allowed investigators to describe the genetic and epigenetic alterations and the mutational spectrum of these tumors. Early reports demonstrated frequent overexpression of the epidermal growth factor receptor (EGFR). Overexpression was shown to correlate with poor prognosis. However, it has not proved to be a good predictor of tumor response to EGFR inhibitors, which are active in only about 10–15% of patients as single agents. Complex genetic analyses, including those by The Cancer Genome Atlas project, have been performed. *p53* mutations are found frequently with other major affected oncogenic driver pathways including the mitotic signaling and Notch pathways and cell cycle regulation in HPV-negative tumors. HPV oncogenes act through direct inhibition of the *p53* and *RB* tumor-suppressor genes, thereby initiating the carcinogenic process. *HRAS* appears to be emerging as a potentially targetable mutation in a small patient subset. While overall mutation rates are similar in HPV-positive and carcinogen-induced tumors, the specific mutational signature of HPV-positive tumors differs, with frequent alteration of the PI3K pathway and occasional mutations in *KRAS*. Overall, these alterations affect mitogenic signaling, genetic stability, cellular proliferation, and differentiation.

CLINICAL PRESENTATION AND DIFFERENTIAL DIAGNOSIS

Most tobacco-related head and neck cancers occur in patients older than age 60 years. HPV-related malignancies are frequently diagnosed in younger patients, usually in their forties or fifties, whereas EBV-related nasopharyngeal cancer can occur at all ages, including in teenagers. The manifestations vary according to the stage and primary site of the tumor. Patients with nonspecific signs and symptoms in the head and neck area should be evaluated with a thorough otolaryngologic examination, particularly if symptoms persist longer than 2–4 weeks. Males are more frequently affected than women by head and neck cancers, including HPV-positive tumors.

Cancer of the nasopharynx typically does not cause early symptoms. However, it may cause unilateral serous otitis media due to obstruction of the eustachian tube, unilateral or bilateral nasal obstruction, or epistaxis. Advanced nasopharyngeal carcinoma causes neuropathies of the cranial nerves due to skull base involvement.

Carcinomas of the oral cavity present as nonhealing ulcers, changes in the fit of dentures, or painful lesions and masses. Tumors of the tongue base or oropharynx can cause decreased tongue mobility and alterations in speech. Cancers of the oropharynx or hypopharynx rarely cause early symptoms, but they may cause sore throat and/or otalgia. HPV-related tumors frequently present with neck lymphadenopathy as the first sign.

Hoarseness may be an early symptom of laryngeal cancer, and persistent hoarseness requires referral to a specialist for indirect laryngoscopy and/or radiographic studies. If a head and neck lesion treated initially with antibiotics does not resolve in a short period, further workup is indicated; to simply continue the antibiotic treatment may be to lose the chance of early diagnosis of a malignancy.

Advanced head and neck cancers in any location can cause severe pain, otalgia, airway obstruction, cranial neuropathies, trismus, odynophagia, dysphagia, decreased tongue mobility, fistulas, skin involvement, and massive cervical lymphadenopathy, which may be unilateral or bilateral. Some patients have enlarged lymph nodes even though no primary lesion can be detected by endoscopy or biopsy; these patients are considered to have

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 77: Head and Neck Cancer, Everett E. Vokes
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

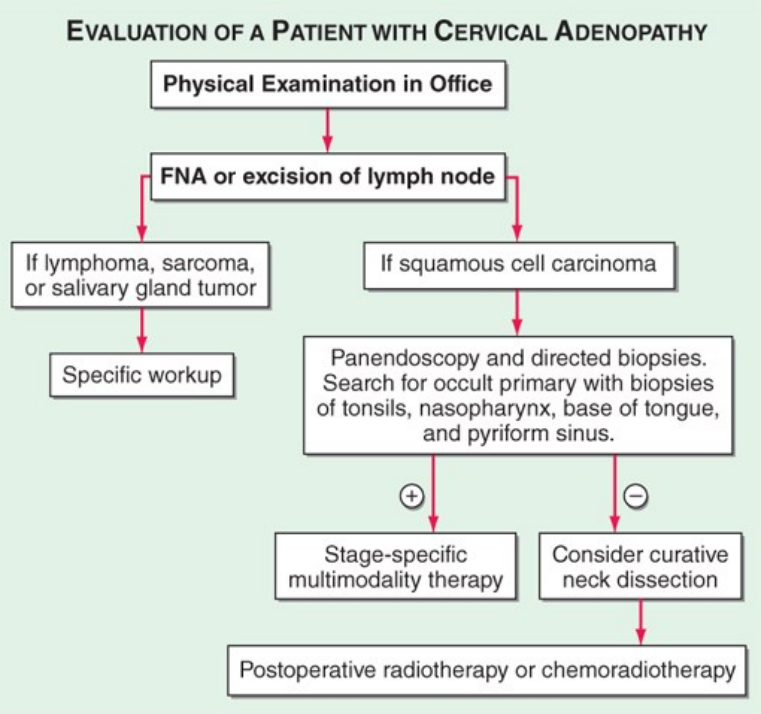
in the chest or abdomen.

Access Provided by:
SILVERCHAIR

patients have enlarged lymph nodes even though no primary lesion can be detected by endoscopy or biopsy; these patients are considered to have carcinoma of unknown primary **(Fig. 77-1)**. Tonsillectomy and directed biopsies of the base of tongue can frequently identify a small primary tumor that frequently will be HPV related. If the enlarged nodes are located in the upper neck and the tumor cells are of squamous cell histology, the malignancy probably arose from a mucosal surface in the head or neck. Tumor cells in supraclavicular lymph nodes may also arise from a primary site in the chest or abdomen.

FIGURE 77-1

Evaluation of a patient with cervical adenopathy without a primary mucosal lesion; a diagnostic workup. FNA, fine-needle aspiration.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

The physical examination should include inspection of all visible mucosal surfaces and palpation of the floor of the mouth and of the tongue and neck. In addition to tumors themselves, leukoplakia (a white mucosal patch) or erythroplakia (a red mucosal patch) may be observed; these "premalignant" lesions can represent hyperplasia, dysplasia, or carcinoma in situ and require biopsy. Further examination should be performed by a specialist. Additional staging procedures include CT or MRI of the head and neck to identify the extent of the disease. Patients with lymph node involvement should have CT scan of the chest and upper abdomen to screen for distant metastases. In heavy smokers, the CT scan of the chest can also serve as a screening tool to rule out a second lung primary tumor. A positron emission tomography (PET) scan can help to identify or exclude distant metastases. CT and PET scans may also be useful in evaluating response to therapy. The definitive staging procedure is an endoscopic examination under anesthesia, which may include laryngoscopy, esophagoscopy, and bronchoscopy; during this procedure, multiple biopsy samples are obtained to establish a primary diagnosis, define the extent of primary disease, and identify any additional premalignant lesions or second primaries.

Head and neck tumors are classified according to the tumor-node-metastasis (TNM) system of the American Joint Committee on Cancer (AJCC) **(Fig. 77-2)**. This classification varies according to the specific anatomic subsite. In general, primary tumors are classified as T1 to T3 by increasing size, whereas T4 usually represents invasion of another structure such as bone, muscle, or root of tongue. Lymph nodes are staged by size, number, and location (ipsilateral vs contralateral to the primary). Distant metastases are found in <10% of patients at initial diagnosis and are more common in patients with advanced lymph node stage; microscopic involvement of the lungs, bones, or liver is more common, particularly in patients with advanced neck lymph node disease. Modern imaging techniques may increase the number of patients with clinically detectable distant metastases in the future. HPV-related oropharyngeal malignancies have consistently been shown to have a better prognosis, and in the eighth edition of the AJCC staging manual, a separate staging system that takes into account the more favorable outlook of these patients will be included. According to this system, patients with advanced nodal stage can still be considered to have an overall early stage (and associated good prognosis), especially if the patient is a nonsmoker or has limited lifelong tobacco exposure.

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 77: Head and Neck Cancer, Everett E. Vokes
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

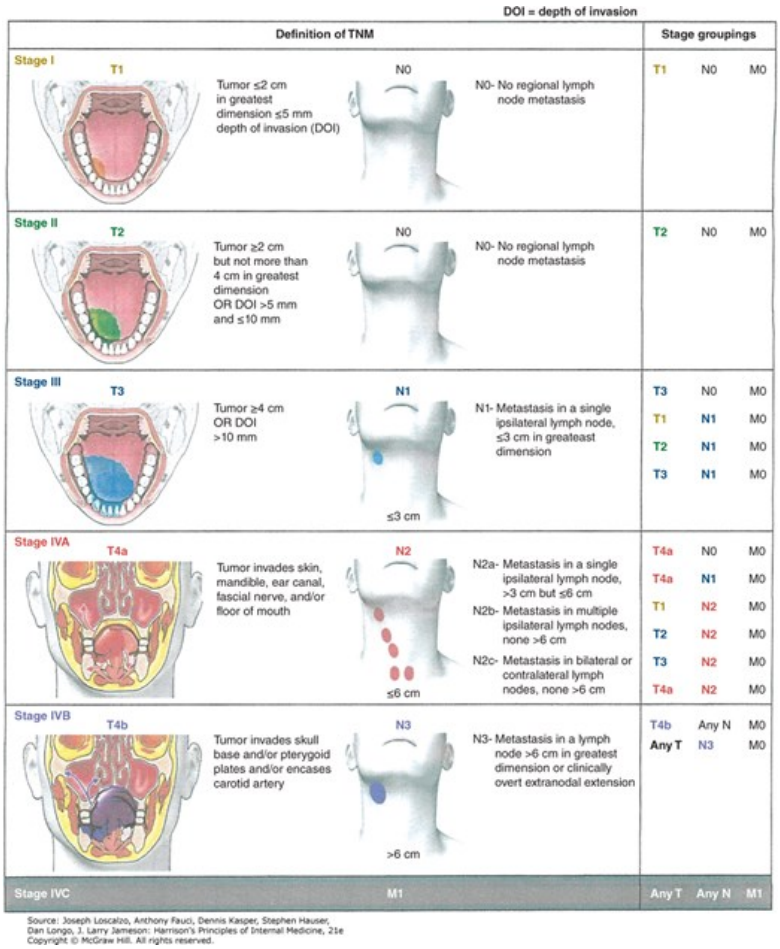
Tumor-node-metastasis (TNM) staging system. *(Figure based on the AJCC Cancer Staging Manual, 8th edition.)*

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION/SYSTEMS

patients with advanced lymph node stage; microscopic involvement of the lungs, bones, or liver is more common, particularly in patients with system, patients with advanced nodal stage can still be considered to have an overall early stage (and associated good prognosis), especially if the patient is a nonsmoker or has limited lifelong tobacco exposure.

FIGURE 77-2

Tumor-node-metastasis (TNM) staging system. *(Figure based on the AJCC Cancer Staging Manual, 8th edition.)*



In patients with lymph node involvement and no visible primary, the diagnosis should be made by lymph node excision (Fig. 77-1). If the results indicate squamous cell carcinoma, a panendoscopy should be performed, with biopsy of all suspicious-appearing areas and directed biopsies of common primary sites, such as the nasopharynx, tonsil, tongue base, and pyriform sinus. HPV-positive tumors especially can have small primary tumors that spread early to locoregional lymph nodes.

TREATMENT

Head and Neck Cancer

Patients with head and neck cancer can be grossly categorized into three clinical groups: those with localized disease, those with locally or regionally advanced disease (lymph node positive), and those with recurrent and/or metastatic disease below the neck. Comorbidities associated with tobacco and alcohol abuse can affect treatment outcome and define long-term risks for patients who are cured of their disease.

LOCALIZED DISEASE

Nearly one-third of patients have localized disease, that is, T1 or T2 (stage I or stage II) lesions without detectable lymph node involvement or distant metastases. These patients are treated with curative intent by either surgery or radiation therapy. The choice of modality differs according to anatomic location and institutional expertise. Radiation therapy is often preferred for laryngeal cancer to preserve voice function, and surgery is preferred for small lesions in the oral cavity to avoid the long-term complications of radiation, such as xerostomia and osteoradionecrosis and dental decay. Randomized data have shown that a prophylactic staging neck dissection should be part of the surgical procedure to eliminate occult

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 77: Head and Neck Cancer, Everett E. Vokes
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Locally or regionally advanced disease—disease with a large primary tumor and/or lymph node metastases—is the stage of presentation for >50% of patients. Such patients can also be treated with curative intent, but not usually with surgery or radiation therapy alone. Combined-modality therapy,

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

preferred for small lesions in the oral cavity to avoid the long-term complications of radiation, such as xerostomia and osteoradionecrosis and dental decay. Randomized data have shown that a prophylactic staging neck dissection should be part of the surgical procedure to eliminate occult nodal metastatic disease. Overall 5-year survival is 60–90%. Most recurrences occur within the first 2 years following diagnosis and are usually local.

LOCALLY OR REGIONALLY ADVANCED DISEASE

Locally or regionally advanced disease—disease with a large primary tumor and/or lymph node metastases—is the stage of presentation for >50% of patients. Such patients can also be treated with curative intent, but not usually with surgery or radiation therapy alone. Combined-modality therapy, including surgery and/or radiation therapy and chemotherapy, is most successful. Chemotherapy can be administered as induction chemotherapy (chemotherapy before surgery and/or radiotherapy) or as concomitant (simultaneous) chemotherapy and radiation therapy. The latter is currently most commonly used and supported by the best evidence. Five-year survival rates exceed 50% in many trials, but part of this increased survival may be due to an increasing fraction of study populations with HPV-related tumors who carry a better prognosis. HPV testing of newly diagnosed tumors should be performed for patients with oropharyngeal tumors at the time of diagnosis. Clinical trials for HPV-related tumors are focused on exploring reductions in treatment intensity, especially radiation dose, in order to ameliorate long-term toxicities (fibrosis, swallowing dysfunction).

In patients with intermediate-stage tumors (stage III and early stage IV), concomitant chemoradiotherapy can be administered as a primary treatment for patients with unresectable disease, to pursue an organ-preserving approach especially for patients with laryngeal cancer (omission of surgery), or in the postoperative setting for smaller resectable tumors with adverse prognostic features.

Induction Chemotherapy

In this strategy, patients receive chemotherapy (current standard is a three-drug regimen of docetaxel, cisplatin, and fluorouracil [5-FU]) before surgery and radiation therapy. Most patients who receive three cycles show tumor reduction, and the response is clinically “complete” in up to half of patients. This “sequential” multimodality therapy allows for organ preservation in patients with laryngeal and hypopharyngeal cancer and results in higher cure rates compared with radiotherapy alone.

Concomitant Chemoradiotherapy

With the concomitant strategy, chemotherapy and radiation therapy are given simultaneously rather than in sequence. Tumor recurrences from head and neck cancer develop most commonly locoregionally (in the head and neck area of the primary and draining lymph nodes). The concomitant approach is aimed at enhancing tumor cell killing by radiation therapy in the presence of chemotherapy (radiation enhancement) and is a conceptually attractive approach for bulky tumors. Toxicity (especially mucositis, grade 3 or 4, in 70–80%) is increased with concomitant chemoradiotherapy. However, meta-analyses of randomized trials document an improvement in 5-year survival of 8% with concomitant chemotherapy and radiation therapy. Cisplatin is preferentially given weekly during a course of daily radiotherapy over a 6- to 7-week course. In addition, concomitant chemoradiotherapy produces better laryngectomy-free survival (organ preservation) than radiation therapy alone in patients with advanced larynx cancer. The use of radiation therapy together with cisplatin produces improved survival in patients with advanced nasopharyngeal cancer. The outcome of HPV-related cancers seems to be especially favorable following cisplatin-based chemoradiotherapy. Trials substituting cisplatin with the EGFR inhibitor cetuximab in that patient population have shown inferior survival.

The success of concomitant chemoradiotherapy in patients with unresectable disease has led to the testing of a similar approach in patients with resected intermediate-stage disease as a postoperative therapy. Concomitant chemoradiotherapy produces a significant improvement over postoperative radiation therapy alone for patients whose tumors demonstrate higher risk features, such as extracapsular spread beyond involved lymph nodes, involvement of multiple lymph nodes, or positive margins at the primary site following surgery.

A monoclonal antibody to EGFR (cetuximab) increases survival rates when administered during radiotherapy. EGFR blockade results in radiation sensitization and has milder systemic side effects than traditional chemotherapy agents, although an acneiform skin rash is commonly observed. Nevertheless, the addition of cetuximab to current standard chemoradiotherapy regimens has failed to show further improvement in survival and is not recommended.

TREATMENT APPROACHES FOR HPV-RELATED HEAD AND NECK CANCERS

Given consistent observations of high survival rates for patients with advanced HPV-related oropharyngeal tumors using combined-modality treatment strategies, de-escalation protocols have attracted widespread interest. The goal here is to decrease the long-term morbidity resulting from high-dose radiation therapy, including extensive neck fibrosis, swallowing problems, and osteoradionecrosis of the jaw. Current studies are investigating the use of lower radiation doses, the use of induction chemotherapy and subsequent omission of chemotherapy or administration of significantly reduced chemoradiation doses in very good responders, and other strategies. In addition, interest has increased in surgical approaches using robotic surgery, which allows better visualization of the base of tongue and tonsil. While technically feasible, this approach tumors will be defined in the coming years.

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 77: Head and Neck Cancer, Everett E. Vokes
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

RECURRENT AND/OR METASTATIC DISEASE

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

significantly reduced chemoradiation doses in very good responders, and other strategies. In addition, interest has increased in surgical approaches using robotic surgery, which allows better visualization of the base of tongue and tonsil. While technically feasible, this approach remains investigational because a large number of patients with disease involving multiple lymph nodes will still require postoperative chemoradiotherapy, thus negating the goal of treatment de-escalation. It is expected that distinct treatment guidelines from carcinogen-induced tumors will be defined in the coming years.

RECURRENT AND/OR METASTATIC DISEASE

Five to ten percent of patients present with metastatic disease, and 30–50% of patients with locoregionally advanced disease experience recurrence, frequently outside the head and neck region. Patients with recurrent and/or metastatic disease are, with few exceptions, treated with palliative intent. Some patients may require local or regional radiation therapy for pain control, but most are given systemic therapy.

Combination chemotherapy formerly was the first-line systemic therapeutic approach to patients with recurrent disease after prior curative intent surgery and/or chemoradiotherapy or those presenting initially with metastatic disease. In particular, a combination of cisplatin with 5-FU and cetuximab (the EXTREME regimen) was frequently used.

However, immunotherapies have proven to be of value in this setting. In particular, inhibitors of the immunosuppressive lymphocyte surface receptor (PD-1) pathway have shown activity in squamous cell cancers of the head and neck. A randomized trial evaluating the PD-1 inhibitor nivolumab versus traditional chemotherapy in the second-line treatment of patients with recurrent or metastatic disease showed a significant increase in 1-year survival rates with fewer severe treatment-related toxicities. In addition, some responses were of long duration, allowing a cohort of patients to live far beyond the historical median of <1 year. The PD-1 inhibitor pembrolizumab also demonstrated activity in a similarly designed randomized trial.

Following establishment of second-line activity, pembrolizumab was compared as single-agent therapy or in combination with cisplatin and 5-FU with prior standard chemotherapy alone (cisplatin, 5-FU, and cetuximab). In this trial, overall survival was improved with pembrolizumab versus chemotherapy as well as with the combination of chemotherapy plus pembrolizumab. No statistically significant impact on progression-free survival was noted. In addition, expression of PD-L1 in the tumor tissue was shown to be of importance. Patients with tumors high in expression (PD-L1 score >20%; i.e., expression of PD-L1 on 20% of tumor cells) had a marked survival benefit with pembrolizumab as single agent, whereas patients with lower PD-L1 expression had a less impressive but still statistically significant survival benefit. However, for the group expressing lower levels of PD-L1, the combination of pembrolizumab with chemotherapy showed more substantial benefit. Current standard treatment therefore frequently consists of combination chemoimmunotherapy for patients with a low combined positive score (CPS; the fraction of tumor cells expressing PD-L1), whereas those with higher CPS scores can be treated with immunotherapy alone, especially if overall tumor burden is limited. Patients who experience progression after first-line chemoimmunotherapy or immunotherapy can then be treated with additional single-agent or combination chemotherapy.

EGFR-directed therapies, including monoclonal antibodies (e.g., cetuximab) and tyrosine kinase inhibitors (TKIs) of the EGFR signaling pathway (e.g., erlotinib or gefitinib), have single-agent activity of ~10%. Side effects are usually limited to an acneiform rash and diarrhea (for the TKIs). The addition of cetuximab to standard combination chemotherapy with cisplatin or carboplatin and 5-FU results in a significant increase in median survival. Drugs targeting specific mutations are under investigation, and patients with *HRAS* mutations have tumor shrinkage with the farnesyltransferase inhibitor tipifarnib.

COMPLICATIONS

Complications from treatment of head and neck cancer are usually correlated to the extent of surgery and exposure of normal tissue structures to radiation. Currently, the extent of surgery has been limited or completely replaced by chemotherapy and radiation therapy as the primary approach. Acute complications of radiation include mucositis and dysphagia. Long-term complications include xerostomia, loss of taste, decreased tongue mobility, second malignancies, dysphagia, and neck fibrosis. The complications of chemotherapy vary with the regimen used but usually include myelosuppression, mucositis, nausea and vomiting, and nephrotoxicity (with cisplatin).

The mucosal side effects of therapy can lead to malnutrition and dehydration. Many centers address issues of dentition before starting treatment, and some place feeding tubes to ensure control of hydration and nutrition intake. About 50% of patients develop hypothyroidism from the treatment; thus, thyroid function should be monitored.

SALIVARY GLAND TUMORS

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 77: Head and Neck Cancer, Everett E. Vokes

©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

may occur as late as 10–20 years after the initial diagnosis. For metastatic disease, therapy is given with palliative intent, usually chemotherapy with

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

SALIVARY GLAND TUMORS

Most benign salivary gland tumors are treated with surgical excision, and patients with invasive salivary gland tumors are treated with surgery and radiation therapy. These tumors may recur regionally; adenoid cystic carcinoma has a tendency to recur along the nerve tracks. Distant metastases may occur as late as 10–20 years after the initial diagnosis. For metastatic disease, therapy is given with palliative intent, usually chemotherapy with doxorubicin and/or cisplatin. Identification of novel agents with activity in these tumors is a high priority. It is hoped that comprehensive genomic characterization of these rare tumors will facilitate these efforts.

FURTHER READING

Agrawal N et al: Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 333:1154, 2011. [PubMed: 21798897]

Burtness B et al: Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): A randomized, open-label, phase 3 study. Lancet 394:1915, 2019. [PubMed: 31679945]

Chow LQM: Head and neck cancer. N Engl J Med 382:60, 2020. [PubMed: 31893516]

D'Cruz AK et al: Elective versus therapeutic neck dissection in node-negative oral cancer. N Engl J Med 373:521, 2015. [PubMed: 26027881]

Ferris RL et al: Nivolumab for recurrent squamous-cell carcinoma of the head and neck. N Engl J Med 375:1856, 2016. [PubMed: 27718784]

Gillison ML et al: Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst 100:407, 2008. [PubMed: 18334711]

Kang H et al: Whole-exome sequencing of salivary gland mucoepidermoid carcinoma. Clin Cancer Res 23:283, 2017. [PubMed: 27340278]

Mehanna H et al: De-escalation after DE-ESCALATE and RTOG 1016: A Head and Neck Cancer Intergroup framework for future de-escalation studies. J Clin Oncol 38:2552, 2020. [PubMed: 32496903]

Sabatini ME et al: Human papillomavirus as a driver of head and neck cancers. Br J Cancer 122:306, 2020. [PubMed: 31708575]

Tota JE et al: Evolution of the oropharynx cancer epidemic in the United States: Moderation of increasing incidence in younger individuals and shift in the burden to older individuals. J Clin Oncol 37:1538, 2019. [PubMed: 31026209]

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 77: Head and Neck Cancer, Everett E. Vokes
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Chapter 85: Renal Cell Carcinoma

Robert J. Motzer; Martin H. Voss

INTRODUCTION

Renal cell carcinomas account for 90–95% of malignant neoplasms arising from the kidney. Notable features include frequent diagnosis without symptoms, resistance to cytotoxic agents, robust activity of angiogenesis-targeted agents, immune infiltration commonly rendering tumors susceptible to checkpoint-directed immunotherapy, and a variable clinical course for patients with metastatic disease, including anecdotal reports of spontaneous regression. Most of the remaining 5–10% of malignant neoplasms arising from the kidney are transitional cell carcinomas (urothelial carcinomas) originating in the lining of the renal pelvis. See **Chap. 86** for transitional cell carcinomas.

EPIDEMIOLOGY

The incidence of cancers of the kidney and renal pelvis rose for three decades, reached a plateau of approximately 64,000 cases annually in the United States between 2012 and 2018, but has since increased to approximately 76,000 cases annually, resulting in close to 14,000 deaths per year. It is the eighth most common cancer overall in the United States, the sixth most common in males, and the eighth most common in females; the male-to-female ratio is 2:1. Although this malignancy may be diagnosed at any age, it is uncommon in those under 45 years, and incidence peaks between the ages of 55 and 75 years. Many factors have been investigated as possible contributing causes; associations include cigarette smoking, obesity, and hypertension. Risk is also increased for patients with polycystic kidney disease that has been complicated by chronic renal failure.

Most cases of renal cell carcinoma (RCC) are sporadic, although familial forms have been reported **(Table 85-1)**. One well established example includes clear cell RCC arising in the context of von Hippel-Lindau (VHL) syndrome, an autosomal dominant disorder. Genetic studies identified the *VHL* gene on the short arm of chromosome 3. Individuals with VHL syndrome have an estimated lifetime risk of developing clear cell RCC of around 70%. Other *VHL*-associated neoplasms include retinal hemangioma, hemangioblastoma of the spinal cord and cerebellum, pheochromocytoma, and neuroendocrine tumors and cysts. Birt-Hogg-Dubé syndrome is a rare human autosomal dominant genetic disorder characterized by fibrofolliculomas (benign tumors arising in hair follicles), pulmonary cysts, and renal cell carcinomas of varying histologies, most commonly the chromophobe type, occurring in about a third of patients. This disorder is associated with mutations in the *FLCN* gene, which codes for folliculin. Other hereditary syndromes are summarized in Table 85-1.

TABLE 85-1

Hereditary Renal Cell Tumors

SYNDROME	CHROMOSOME(S)	GENE	PROTEIN	KIDNEY TUMOR TYPE	ADDITIONAL CLINICAL FINDINGS
von Hippel-Lindau syndrome	3p25	*VHL*	von Hippel-Lindau protein	Clear cell	Hemangioblastoma of the retina and central nervous system; pheochromocytoma; pancreatic and renal cysts; neuroendocrine tumors
Hereditary papillary RCC	7p31	*MET*	MET	Papillary (type I)	Bilateral and multifocal kidney tumors
Hereditary leiomyomatosis	1q42	*FH*	Fumarate hydratase	Papillary (non–type I)	Leiomyoma; uterine leiomyoma/leiomyosarcoma
Birt-Hogg-Dubé syndrome	17p11	*FLCN*	Folliculin	Chromophobe; oncocytoma	Facial fibrofolliculoma; pulmonary cysts

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 85: Renal Cell Carcinoma, Robert J. Motzer; Martin H. Voss
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®
Access Provided by:
SILVERCHAIR INFORMATION / SYSTEMS

syndromes are summarized in Table 85-1.

TABLE 85-1

Hereditary Renal Cell Tumors

SYNDROME	CHROMOSOME(S)	GENE	PROTEIN	KIDNEY TUMOR TYPE	ADDITIONAL CLINICAL FINDINGS
von Hippel-Lindau syndrome	3p25	*VHL*	von Hippel-Lindau protein	Clear cell	Hemangioblastoma of the retina and central nervous system; pheochromocytoma; pancreatic and renal cysts; neuroendocrine tumors
Hereditary papillary RCC	7p31	*MET*	MET	Papillary (type I)	Bilateral and multifocal kidney tumors
Hereditary leiomyomatosis and RCC (HLRCC syndrome)	1q42	*FH*	Fumarate hydratase	Papillary (non–type I)	Leiomyoma; uterine leiomyoma/leiomyosarcoma
Birt-Hogg-Dubé syndrome	17p11	*FLCN*	Folliculin	Chromophobe; oncocytoma	Facial fibrofolliculoma; pulmonary cysts
Tuberous sclerosis	9q34 16p13	*TSC1* *TSC2*	Hamartin Tuberin	Angiomyolipomas; lymphangioleiomyomatosis; rare RCC with variety of histologic appearances	Angiofibroma, subungual fibroma; cardiac rhabdomyoma; adenomatous small intestine polyps; pulmonary and renal cysts; cortical tuber; subependymal giant cell astrocytomas
BAP1 tumor predisposition syndrome	3p21	*BAP1*	BAP1	Clear cell	Atypical Spitz tumors; uveal melanoma; cutaneous melanoma; basal cell carcinoma; malignant mesothelioma

Abbreviation: RCC, renal cell carcinoma.

PATHOLOGY AND GENETICS

Renal cell malignancies represent a heterogeneous group of tumors with distinct histopathologic, genetic, and clinical features (Table 85-2). Categories include clear cell carcinoma (70% of cases), papillary tumors (10–15%), chromophobe tumors (≤5%), renal medullary carcinoma (<1%), translocation carcinoma (<5%), and other less common variants. Papillary tumors can be bilateral and multifocal. Chromophobe tumors tend to have a more indolent clinical course. Translocation-associated RCC, rare in adult patients, is the predominant histology in children. Renal medullary carcinoma is rare, very aggressive, and associated with sickle cell trait. Tumors that do not meet criteria for defined variants are generally referred to as "unclassified" with variable clinical courses.

TABLE 85-2

Classification of Malignant Epithelial Neoplasms Arising from the Kidney

CARCINOMA TYPE	CHARACTERISTIC GROWTH PATTERN	CHROMOSOMAL EVENTS	GENES WITH RECURRENT SOMATIC ALTERATIONS
Clear cell	Acinar or sarcomatoid	3p–, 5q+, 14q–, 9p–	*VHL, PBRM1, BAP1, SETD2*
Chromophobe	Solid, tubular, or sarcomatoid	Whole arm losses (1, 2, 6, 10, 13, 17, and 21)	*TP53, PTEN, TERT* promoter

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 85: Renal Cell Carcinoma, Robert J. Motzer; Martin H. Voss
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

TABLE 85-2

Classification of Malignant Epithelial Neoplasms Arising from the Kidney

CARCINOMA TYPE	CHARACTERISTIC GROWTH PATTERN	CHROMOSOMAL EVENTS	GENES WITH RECURRENT SOMATIC ALTERATIONS
Clear cell	Acinar or sarcomatoid	3p–, 5q+, 14q–, 9p–	*VHL, PBRM1, BAP1, SETD2*
Papillary	Papillary or sarcomatoid	+7, +17, 9p–	*MET, FH, CDKN2A* (focal deletions)
Chromophobe	Solid, tubular, or sarcomatoid	Whole arm losses (1, 2, 6, 10, 13, 17, and 21)	*TP53*, *PTEN*, *TERT* promotor
Renal medullary carcinoma	Varying growth patterns, including cribriform, reticular, sarcomatoid, adenoid, and microcystic	+8q, 22q–, 22q translocations	*SMARCB1* (focal deletions, mutations, gene fusions), *SETD2*
MITF translocationa	Mimicking clear cell and papillary variants	Xp11.2 translocations; t(6;11) translocations	*TFE3* gene fusions, *TFEB* gene fusions

[a]Microphthalmia transcription factor gene family.

Clear cell tumors, the predominant histology, are found in >80% of patients who develop metastases and arise from the epithelial cells of the proximal tubules. Loss of chromosome 3p is uniformly seen as the earliest event in the development of these cancers. This leads to loss of heterozygosity for a number of relevant 3p genes, including *VHL*, *PBRM1*, *BAP1*, and *SETD2*, which can be functionally lost through secondary events in the remaining allele. *VHL* encodes a tumor suppressor protein that is involved in regulating the transcription of vascular endothelial growth factor (VEGF) and a number of other effectors through ubiquitination of hypoxia-inducible factors (HIF). Inactivation of *VHL*, through upregulation of VEGF signaling, promotes tumor angiogenesis and growth, ultimately rendering clear cell RCC cells susceptible to antiangiogenesis therapy.

Large-scale sequencing efforts have helped elucidate recurrent patterns of genomic evolution that correlate with distinct clinical phenotypes, e.g., varying levels of aggressiveness or specific patterns of metastatic spread. For example, early loss of chromosome 9p appears to confer a high risk for early metastatic dissemination and correlates with poor cancer-specific survival.

A growing number of other RCC variants are well defined (see Table 85-2 for examples). For instance, up to 15% of RCCs are of the papillary subtype, with several subtypes that can be distinguished either by light microscopy or tumor genomics. For example, activating mutations in the *MET* oncogene or gain of chromosome 7 (where *MET* is located) are hallmark events of type 1 papillary RCC and considered actionable via targeted MET inhibitors. Tumors of the less common chromophobe subtype originate from the distal nephron. They are in part driven by changes in mitochondrial gene function and typically characterized by aneuploidy with common loss of an entire chromosome copy for chromosomes 1, 2, 6, 10, 13, and 17.

CLINICAL PRESENTATION

Presenting signs and symptoms may include hematuria, flank or abdominal pain, and a palpable mass. Other symptoms are fever, weight loss, anemia, and a varicocele. Tumors are, however, commonly detected as an incidental finding on a radiograph. Widespread use of radiologic cross-sectional imaging (computed tomography [CT], magnetic resonance imaging [MRI]) contributes to earlier detection of renal masses during evaluation for other medical conditions. The increasing number of incidentally discovered low-stage tumors has contributed to an improved 5-year survival for patients with RCC and increased use of nephron-sparing surgery (partial nephrectomy). A spectrum of paraneoplastic syndromes has been associated with these malignancies, including erythrocytosis, hypercalcemia, nonmetastatic hepatic dysfunction (Stauffer's syndrome), and acquired dysfibrinogenemia. Erythrocytosis is noted at presentation in only about 3% of patients. Anemia, commonly a sign of more advanced disease, is more common. Kidney cancer was called the "internist's tumor" since it was often discovered from the initial presentation of a paraneoplastic syndrome. This was more common before the era of modern imaging, as was initial presentation by the classic triad of hematuria, flank pain, and a palpable abdominal mass.

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 85: Renal Cell Carcinoma, Robert J. Motzer; Martin H. Voss
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

suspected tumor involvement or invasion by thrombus, or when intravenous contrast administration given with CT is prohibited by impaired renal function. In clinical practice, any solid renal masses should be considered malignant until proven otherwise; a definitive diagnosis is required. If no

This was more common before the era of modern imaging, as was initial presentation by the classic triad of hematuria, flank pain, and a palpable abdominal mass.

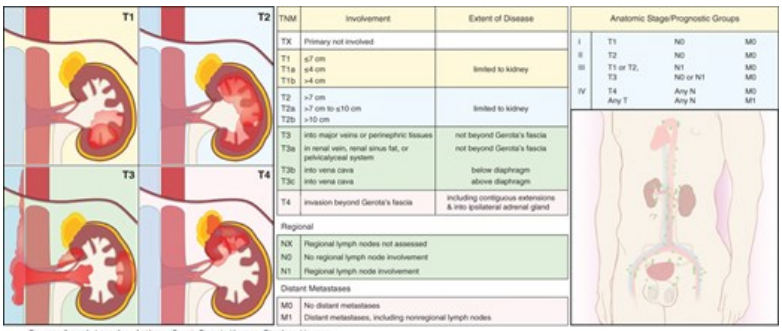
The standard evaluation of patients with suspected renal tumors includes a CT scan of the abdomen and pelvis, chest radiograph, and urine analysis. If metastatic disease is suspected from the chest radiograph, a CT of the chest is warranted. MRI is useful in evaluating the inferior vena cava in cases of suspected tumor involvement or invasion by thrombus, or when intravenous contrast administration given with CT is prohibited by impaired renal function. In clinical practice, any solid renal masses should be considered malignant until proven otherwise; a definitive diagnosis is required. If no metastases are demonstrated, surgery is indicated, even if the renal vein or inferior vena cava is invaded. In small tumors (particularly those of clear cell variant) the risk of impending metastatic spread is lower and surgery can potentially be delayed. In that setting, a needle biopsy should be performed to confirm the underlying histology, and radiographic surveillance is indicated until the time of surgery. The differential diagnosis of a renal mass includes cysts, benign neoplasms (adenoma, angiomyolipoma, oncocytoma), inflammatory lesions (pyelonephritis or abscesses), and other malignancies originating in the kidney such as transitional cell carcinoma of the renal pelvis, sarcoma, lymphoma, and Wilms' tumor or metastases from cancers originating in other organs. All of these are less common causes of renal masses than is RCC. The most common sites of distant metastases are the lungs, lymph nodes, liver, bone, and brain. These tumors may follow an unpredictable and protracted clinical course.

STAGING AND PROGNOSIS

Staging is based on the American Joint Committee on Cancer (AJCC) staging system **(Fig. 85-1)**. Stage I tumors are ≤7 cm in greatest diameter and confined to the kidney, stage II tumors are >7 cm and confined to the kidney, stage III tumors extend through the renal capsule but are confined to Gerota's fascia, grossly infiltrate the renal vein, or involve regional lymph nodes (N1), and stage IV disease includes tumors that have invaded adjacent organs or involve nonregional lymph nodes or distant metastases. Sixty-five percent of patients present with stage I or II disease, 15–20% with stage III, and 15–20% with stage IV. The 5-year survival rate is currently 75% across all RCCs, but varies greatly by stage.

FIGURE 85-1

Renal cell carcinoma staging. TNM, tumor-node-metastasis.



Source: Joseph Loscalzo, Anthony Fauci, Dennis Kasper, Stephen Hauser, Dan Longo, J. Larry Jameson: Harrison's Principles of Internal Medicine, 21e
Copyright © McGraw Hill. All rights reserved.

Prognostic risk models are helpful for counseling patients diagnosed with metastatic disease and for anticipating survival rates when designing a clinical trial. A widely used prognostic model for advanced disease, the International Metastatic RCC Database Consortium (IMDC) risk model, incorporates six factors shown to correlate with worse survival: poor performance status, low hemoglobin concentration, high serum calcium, high neutrophil levels, high platelet levels, and <1-year interval from diagnosis to systemic treatment. Patients with zero risk factors had significantly longer median survival (43 months) than patients with one or two risk factors (22.5 months) and those with three to six risk factors (8 months) when treated with first-line angiogenesis inhibitors (see below).

TREATMENT

Renal Cell Carcinoma

LOCALIZED TUMOR

The standard management for stage I or II tumors and selected cases of stage III disease is radical or partial nephrectomy. A radical nephrectomy involves en bloc removal of Gerota's fascia and its contents, including the kidney, and commonly the ipsilateral adrenal gland and regional lymph nodes that appear abnormal on imaging or intraoperatively. Open, laparoscopic, or robotic surgical techniques may be used. The role of a template

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 85: Renal Cell Carcinoma, Robert J. Motzer; Martin H. Voss
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

Nephron-sparing approaches, i.e., open or laparoscopic partial nephrectomy, may be appropriate depending on the size and location of the tumor. This approach is particularly relevant for patients with solitary kidneys, bilateral tumors, or chronic renal insufficiency but can also be applied

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

nodes that appear abnormal on imaging or intraoperatively. Open, laparoscopic, or robotic surgical techniques may be used. The role of a template lymphadenectomy in patients without apparent lymphadenopathy is controversial. Extension into the renal vein or inferior vena cava (stage III disease) does not preclude resection, which would then include thrombectomy.

Nephron-sparing approaches, i.e., open or laparoscopic partial nephrectomy, may be appropriate depending on the size and location of the tumor. This approach is particularly relevant for patients with solitary kidneys, bilateral tumors, or chronic renal insufficiency but can also be applied electively to resect small masses for patients with normal kidney function. Radical nephrectomy carries a greater risk for chronic kidney disease and cardiovascular morbidity and mortality.

Adjuvant systemic therapies, including cytokines and targeted agents, have been studied in randomized clinical trials, largely with negative results, and the standard of care remains active surveillance after nephrectomy.

METASTATIC DISEASE

Surgery has a limited role for patients with metastatic disease. Long-term survival may occur in patients who relapse with a solitary site that is removed (metastasectomy). Nephrectomy despite presence of metastases (cytoreductive nephrectomy) is considered for carefully selected patients with stage IV disease. One indication for this approach can be to alleviate pain or hemorrhage of a primary tumor.

Radiation therapy is used for palliation of bone or brain metastases. The type of radiotherapy most commonly used is external-beam therapy, including stereotactic radiosurgery and other forms of image-guided radiotherapy.

Systemic therapy is the mainstay of care for metastatic disease. The timing of initiating such treatment should be carefully considered; some patients are asymptomatic at diagnosis, and with indolent behavior, it may be best to document progression before initiating treatment.

Metastatic RCC is refractory to cytotoxic chemotherapy. Patients are treated with molecularly targeted agents, including targeted immunotherapies. Treatments are continued with noncurative intent while tolerated and until disease progression is evident on cross-sectional imaging. Outcomes for patients with metastatic disease improved when increased understanding of underlying biology led to the successful development of several tyrosine kinase inhibitors (TKIs) targeting proangiogenic signaling through the VEGF receptors as well as allosteric inhibitors of mammalian target of rapamycin (mTOR) signaling. Serial large-scale randomized trials demonstrated that such agents, typically orally available, could be administered sequentially and in combination. Pivotal studies, by design, defined a dedicated space for each regimen in treatment-naïve or pretreated patients **(Table 85-3)**.

Targeted immunotherapies were introduced after VEGF- and mTOR-directed agents had established standards of care in the first- and second-line setting. Nivolumab, a checkpoint inhibitor targeting PD-1, was compared to the mTOR inhibitor everolimus in a randomized trial in patients who had progressed on prior TKI therapy, challenging the standard approach in pretreated patients. Nivolumab demonstrated superior overall survival, positioning it as the new second-line agent of choice. Subsequently, immunotherapy combination regimens demonstrated efficacy in randomized trials conducted in treatment-naïve patients. In separate studies, two doublets demonstrated survival benefit over standard sunitinib therapy and changed the standard of care for untreated metastatic clear cell RCC: nivolumab in combination with the CTLA-4-directed checkpoint inhibitor ipilimumab proved superior to sunitinib in patients with high-risk features per the IMDC model, achieving complete radiographic disappearance of cancer in >10% of patients treated with the combination. In a second trial, the combination of the TKI axitinib together with the PD-1 inhibitor pembrolizumab was superior to sunitinib alone in all-comers with untreated metastatic RCC, again with high response rates across all IMDC risk groups. In both trials, responses were long-lasting, with improved time to disease progression and longer overall survival for combination regimens. Additional trials are ongoing to fortify the new standard of combination therapy in the first line.

With an ever-growing number of approved options directed toward different molecular targets, biomarkers are urgently needed to help individualize therapeutic choices and to gain insight as to whether and why treatments are working. Although a multitude of candidate biomarkers have been investigated for their predictive value in metastatic RCC, none have been validated for clinical use to date.

Projected overall survival in patients starting systemic therapies for newly diagnosed metastatic disease has tripled over the past 15–20 years; this can largely be attributed to the successful drug developments discussed here.

TABLE 85-3

Commonly Used Systemic Regimens for Metastatic Renal Cell Carcinoma

Antiangiogenic: TKIs	Sunitinib	2006	Advanced RCC, first line

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 85: Renal Cell Carcinoma, Robert J. Motzer; Martin H. Voss
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

TABLE 85-3

Commonly Used Systemic Regimens for Metastatic Renal Cell Carcinoma

CLASS	DRUG	FIRST FDA APPROVAL FOR RCC	CURRENTLY USED FOR
Antiangiogenic: TKIs	Sunitinib	2006	Advanced RCC, first line
	Pazopanib	2009	Advanced RCC, first line
	Axitinib	2012	Advanced RCC, pretreated
	Cabozantinib	2016 2017	Advanced RCC, pretreated with antiangiogenic therapy Advanced RCC, first line
Immunotherapy: checkpoint inhibitor	Nivolumab	2015	Advanced RCC, pretreated with antiangiogenic therapy
Combination therapies			
TKI + mTOR inhibitor	Lenvatinib + everolimus	2016	Advanced RCC, pretreated with one antiangiogenic therapy
PD-1 inhibitor + CTLA-4 inhibitor	Nivolumab + ipilimumab	2018	Advanced intermediate- or poor-risk RCC, first line
PD-1 inhibitor + TKI	Pembrolizumab + axitinib	2019	Advanced RCC, first line

Abbreviations: CTLA-4, cytotoxic T-lymphocyte-associated protein; FDA, U.S. Food and Drug Administration; mTOR, mammalian target of rapamycin; PD-1, programmed cell death-1; RCC, renal cell carcinoma; TKI, tyrosine kinase inhibitor.

GLOBAL CONSIDERATIONS

Worldwide, over 400,000 patients are diagnosed each year with malignant tumors arising from the kidney, resulting in >175,000 deaths annually. Kidney cancer is the 10th most common cancer in men and the 15th most common cancer in women. Higher incidence is observed in developed countries, including the United States, Canada, Europe, Australia, New Zealand, and Uruguay. Relatively low rates are reported in Southeast Asia and Africa. The incidence of kidney cancer has been steadily increasing over the past four decades. Mortality trends have stabilized in Europe and the United States, but not in less developed countries. This is likely related to differences in access to optimal therapies. Treatment guidelines for both localized and metastatic renal cancer are similar between U.S. and European documents and contingent on the access to adequate health care and availability of targeted drugs to treat metastases.

FURTHER READING

Choueiri TK, Motzer RJ: Systemic therapy for metastatic renal cell carcinoma. N Engl J Med 376:354, 2017. [PubMed: 28121507]

Moch H et al: The 2016 WHO classification of tumours of the urinary system and male genital organs—Part A: renal, penile, and testicular tumours. Eur

Motzer RJ et al: Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade. Cancer Cell 38:803, 2020. [PubMed:

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 85: Renal Cell Carcinoma, Robert J. Motzer; Martin H. Voss
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility

ACCESS Medicine®

Access Provided by:
SILVERCHAIR INFORMATION SYSTEMS

Moch H et al: The 2016 WHO classification of tumours of the urinary system and male genital organs—Part A: renal, penile, and testicular tumours. Eur Urol 70:93, 2016. [PubMed: 26935559]

Motzer RJ et al: Nivolumab plus ipilimumab versus sunitinib in advanced renal-cell carcinoma. N Engl J Med 378:1277, 2018. [PubMed: 29562145]

Motzer RJ et al: Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade. Cancer Cell 38:803, 2020. [PubMed: 33157048]

Rini BI et al: Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. N Engl J Med 380:1116, 2019. [PubMed: 30779529]

Sanchez A et al: Current management of small renal masses, including patient selection, renal tumor biopsy, active surveillance, and thermal ablation. J Clin Oncol 36:3591, 2018. [PubMed: 30372390]

Turajlic S et al: Tracking cancer evolution reveals constrained routes to metastases: TRACERx Renal. Cell 173:581, 2018. [PubMed: 29656895]

Wong MCS et al: Incidence and mortality of kidney cancer: Temporal patterns and global trends in 39 countries. Sci Rep 7:15698, 2017. [PubMed: 29146923]

Downloaded 2022-1-3 5:52 P Your IP is 141.216.78.40
Chapter 85: Renal Cell Carcinoma, Robert J. Motzer; Martin H. Voss
©2022 McGraw Hill. All Rights Reserved. Terms of Use • Privacy Policy • Notice • Accessibility